

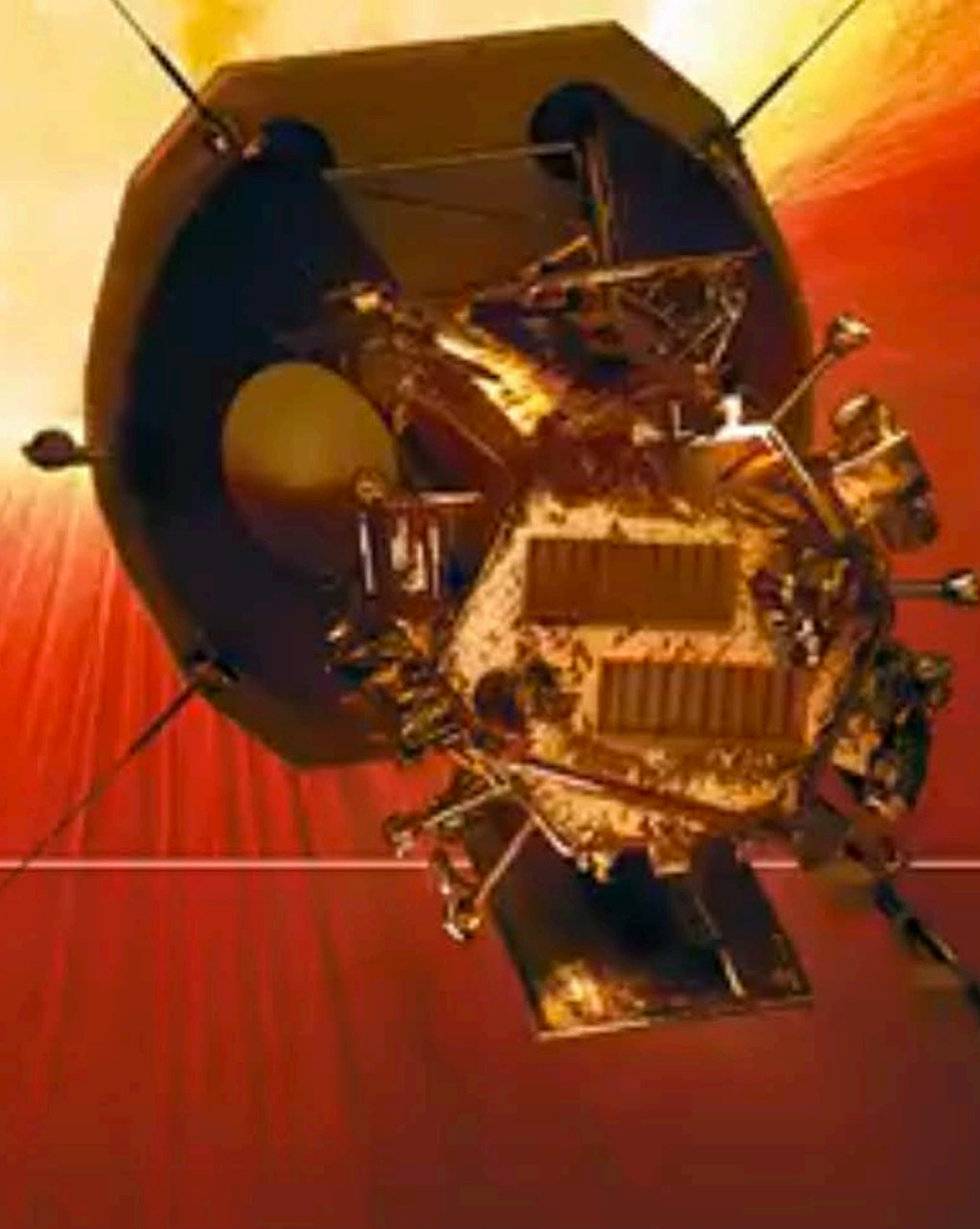
The international journal of science / 12 December 2019

index
Materials
science

nature

STARING AT THE SUN

First results from NASA's
Parker Solar Probe



Predatory journals
One definition to help
tackle unscrupulous
publishers

Mutated proteins
Structural secrets of
drug resistance in
malaria parasites

Patched up
Mobile cellular sealant
helps to heal deep
wounds in the skin

Vol. 576, No. 7786
nature.com

Good practice from the grass roots

Community-led efforts – not just global ones – are key to research integrity.

When it comes to research integrity, scientists use the language of aspiration, whereas policymakers talk about hard rules and enforcement.

That's one conclusion from an in-depth analysis of published research and policy documents in research integrity (S. P. J. M. Horbach and W. Halffman *Sci. Eng. Ethics* **23**, 1461–1485; 2017). There are other disconnects, too. Countries, disciplines and sectors often approach integrity in different ways. For some, it can be confined to preventing data fabrication, falsification and plagiarism. But integrity is much broader, encompassing quality and relevance, as well as recognition of diversity and inclusion.

The need for a unified approach is slowly gaining recognition. The World Science Forum, a biennial meeting of researchers and policymakers from different countries, issued a declaration at its November conference in Budapest that called for, among other things, “harmonisation and enforcement of standards of conduct of scientific research across borders and across public and private research”. The declaration also supported processes by which scientists “can report suspected research misconduct and other irresponsible research practices, without fear of reprisal”, and it urged clearer procedures for responding to such concerns.

These proposals echo many national and international guidance documents that have been produced since 2010, when the Singapore Statement on Research Integrity billed itself as the first international effort to unify approaches to integrity. Subsequent guidelines have included the Bonn PRINTEGER Statement, the All European Academies Code of Conduct for Research Integrity, the Netherlands Code of Conduct for Research Integrity and many more.

But none yet has the globally respected status of, say, the 1964 Declaration of Helsinki, which is widely accepted as the standard for the ethical treatment of human participants in medical research. And that raises the question: how can the burgeoning official declarations actually enhance research integrity in the lab?

For the working researcher, complying with a research-integrity policy can sometimes feel like another unwelcome burden – yet another form to fill in, set of data to record or online system to feed. And because these policies are often handed down from the directors, that can make workers unwilling to accept that these practices are something that everyone benefits from and needs to engage with.

That's where local, community-driven efforts are offering a way forward. On page 183, Marcus Munafò at

“Research-integrity policy can sometimes feel like another unwelcome burden.”

the University of Bristol, UK, describes how one network of universities and journal publishers is enabling members to exchange ideas on integrity, which could lead to consensus about practices such as promoting transparency and data sharing. And on page 210, a multinational team of researchers and publishers presents a consensus definition of predatory journals. This is a welcome and genuinely collaborative effort, the result of many rounds of discussion between researchers, institutional leaders and patient representatives, among other stakeholders.

Networks for integrity are emerging in the global south, too, such as the African Research Integrity Network, which was established in 2017, and the Asia Pacific Research Integrity Network, set up two years earlier, in 2015.

Perhaps some of the most exciting grass-roots efforts are those that connect researchers to each other and to policymakers. The Embassy of Good Science is one such platform. Established under the European Union's Horizon 2020 research programme, it styles itself as a ‘public square’ where researchers can find guidance and share knowledge. It contains links to national ethics codes and articles explaining good practice.

Official declarations about research integrity can sometimes be perceived by researchers as statements of the obvious, particularly because researchers regard the pursuit of science itself as upholding values of fairness, honesty and scepticism. Policy documents, moreover, can be viewed warily by researchers as a list of commandments handed down from on high, chiselled into stone tablets.

But rules and policies are needed – and they can be most effective when they arise as a result of engagement with grass-roots communities. Local efforts might not produce the global unity and consistency of high-level statements, but they have a much higher chance of changing researchers' daily practices, and thus making a genuine difference to integrity in research. The lesson here is familiar: change will come when we work alongside the communities we wish to change.

Space for dialogue

Sustainability calls for collaboration between architects and behavioural scientists.

There was a time when ‘sustainable development’ meant economic development, or perpetual economic growth – not, as we know it today, environmentally sustainable development.

The change in meaning can be traced to the 1987 report *Our Common Future*, chaired by Norway's then-prime minister, Gro Harlem Brundtland. The report involved social scientists, natural scientists, industrialists, environmentalists and policymakers emerging from their silos to talk to each other to understand how humans alter the global environment. The report helped such collaborative

processes to become mainstream, alongside the idea of treating the environment and development as one issue.

Some fields quickly grasped that interdisciplinary work is essential to understanding environmental change, and to mitigating – or adapting to – its effects. Confirming a human cause for climate change required the combined efforts of meteorologists, oceanographers and geographers, among others. Replacing the ozone-depleting chemicals used in spray cans and refrigerators needed chemists to talk to product designers. But, as a report this week in *Nature Sustainability* shows, other fields have not got so far in their interdisciplinary journey (L. Kotz *et al. Nature Sustain.* 2, 1067–1069; 2019).

In a project convened by the journal and the Convergent Behavioral Science Initiative at the University of Virginia in Charlottesville, an international group of architects, designers and engineers spent a year with behavioural scientists, investigating how their disciplines could better work together, and why they needed to do so.

Behavioural science has an existing and essential relationship to the built environment: we have to study how people live, work and move to create liveable buildings and towns. But the group established that, when it comes to sustainability, there's room for closer working, and the report amounts to an agenda for joint research. Potential questions include: how do architects and designers make decisions? To what extent can behavioural science in other contexts be applied to sustainable design and architecture? Do architects feel a duty to promote responsible energy use?

Cross-disciplinary working requires careful communication and confidence-building. As the example of defining sustainable development shows, disciplines have their own languages and can interpret terms differently.

Lessons in interdisciplinarity can also be learnt from the 'science wars' of the mid-1990s, a tense time in the relationship between natural scientists and the sociologists who study how research is done. Part of the ambition for sociologists of science is to place a mirror before researchers, to demonstrate potential flaws in their methods. But some eminent researchers saw these studies as an intrusion, and thought that natural scientists had little to learn from them.

One way to ease disciplinary tensions could be to underscore that sustainability calls for behavioural change at all levels – necessitating more research across all sectors. Governments, for example, often interact with independent researchers who study how to improve policy, including how government itself needs to adapt if it is to drive sustainability more effectively. Similarly, business schools produce case studies on how companies can adapt to facilitate that change. Behavioural research could help all of us – individuals and communities – to make changes to how we behave, whether it is taking more public transport or just turning the thermostat down a degree.

Along with governments, industry and individuals, the built environment consumes energy and produces waste, which makes it just as pivotal to sustainability. As the *Nature Sustainability* report says, collaborating effectively and learning from each other can be tough. But considering the planetary situation, not doing so has much higher costs.

“This week, we begin a more concerted push to promote diversity across our content.”

No more 'manels'

Nature's new code of conduct strives for more diversity at research meetings and events.

That women from under-represented minorities receive few speaking invitations to the world's largest Earth-science conference has again shone a spotlight on science's diversity deficit (H. L. Ford *et al. Nature* 576, 32–35; 2019).

Conferences are essential for research communication, and taking part is important for career progression. But turning the dial on diversity – and stopping it from slipping back – is proving difficult. Our investigation this year of 'manels' and 'manferences' – panels and conferences dominated by male speakers – showed that sometimes a heroic effort to diversify them one year is followed by business as usual the next (*Nature* 573, 184–186; 2019).

At *Nature*, we are aware of our own shortcomings – that our authors and referees, for example, include too few women – and of our responsibilities to turn things around (*Nature* 558, 344; 2018). This week, we begin a more concerted push to promote diversity across our editorial and publishing activities, including concrete commitments in the events that we organize (see go.nature.com/36jtrf).

In 2019, *Nature* and other journals in the Nature Research portfolio hosted, or co-hosted, more than 30 events in a range of disciplines. But despite informal efforts to make our conferences more inclusive, women and people from minority groups still make up only a small proportion of our speakers. We are therefore formalizing our efforts into a published code of conduct. This will apply not only to Nature Conferences but to all scholarly events organized or co-organized by Springer Nature.

The code commits us to having no male-only organizing committees for Nature Conferences planned from this point. We will invite equal numbers of women and men as speakers, whether we're selecting for keynote presentations or from abstract submissions. We also commit to having no manels at our events, and to monitor and report progress against these goals at the end of each calendar year. Planning for most of our events in 2020 is already advanced, so the full effect of our commitment will be seen from 2021.

Nature Conferences must be welcoming, safe, collaborative and productive for all attendees. Our code states that we expect participants to be considerate of diverse views and cultures, and respectful and collaborative in their discussion and critiques of ideas. Appropriate sanctions will be applied where the code is not followed.

We also commit to supporting diversity more broadly, including in geography, ethnicity, culture, career stage, disability and sexual orientation. With time, we aim to develop our code further to address this explicitly.

Scientific events must be more inclusive. We hope that this initiative – like similar ones in many other organizations – goes some way to reaching that goal.

World view



By Marcus
Munafò

Raising research quality will require collective action

Institutions must act together to reform research culture, says Marcus Munafò.

Modern research is far removed from that done by the independent scientist of Charles Darwin's day. But that model still underpins what we reward. And the tension is showing. The number of authors listed on published papers has increased fivefold. But our system of incentives – accreted haphazardly as science evolved from hobby to career – still focuses on individuals.

Funding, appointments, promotions, tenure, prizes and so on emphasize individual achievement and overlook deeds that benefit everyone and should be valued explicitly, such as producing usable tools or sharing code. If we want to move towards a transparent model of research, we need to reward open-research practices. If we want researchers to work well in large collaborations, we need to train them in communication skills and collective self-scrutiny.

In the past five years, many studies have attempted to assess the prevalence of questionable research practices. The Netherlands has commissioned a country-wide survey. Others have focused on specific fields, from ecology and evolution to health services. Describing the scale of the problem is necessary, but insufficient.

We must reflect on how (and, more importantly, why) questionable research practices and undesirable behaviours arise and persist. What are the flaws in our institutions' cultures and practices that allow this conduct to proliferate? We must find the root causes. Have we disincentivized solid, cumulative work, replication studies and publishing null results – essential if science is to self-correct efficiently – with our relentless focus on groundbreaking findings?

That self-inspection is in the air. Institutions are committing to working together to determine how their cultural practices, such as emphasizing the importance of novelty, discovery and priority, undermine the value of replication, verification and transparency. That is the goal of the UK Reproducibility Network, which I co-founded earlier this year. It started as informal groups of researchers at individual institutions that met with representatives from funders and publishers (including *Nature*) who were open to discussions about how best to align open-science initiatives – reproducibility sections in grant applications and reporting checklists in article submissions, for example. Now institutions themselves are cooperating to consider larger changes, from training to hiring and promotion practices.

Our ten university members span the United Kingdom from Aberdeen to Surrey, and we expect that list to grow. Each will appoint a senior academic to focus on research quality and improvement. Figuring out which system-level



If we want to move towards a transparent model of research, we need to reward open-research practices."

Marcus Munafò
is professor of biological psychology at the University of Bristol, UK, and chair of the UK Reproducibility Network Steering Group.
e-mail: marcus.munafò@bristol.ac.uk

changes are needed and how to make them happen will now be someone's primary responsibility, not a volunteer activity. What changes might ensue? Earlier this year, the University of Bristol, where I work, made the use of data sharing and other open-research practices an explicit criterion for promotion.

But one institution will make little difference on its own. For better practices to become the norm, many universities need to act collectively. Changes to incentives at a single institution will not make new behaviours stick, not least because practices required in only one place can act as a career tax on its scientists. Only if changes occur across many institutions will the impacts permeate scientific culture.

The same is true for training. If universities agree that all their graduates must reach common standards – in data skills, for example – future scientists will do better research and collaborate across institutions in a way that is hard to imagine today. Consider the productivity gains if competence in the programming language R – which can be used to run statistical analyses across a broad range of areas – was a given. Research would become more efficient, like the railways did after adopting a common standard for track gauge.

Public-health campaigns to reduce smoking demonstrate how coordinated action can yield broad cultural change. In 1974, 45% of the UK population smoked; now it is below 15%. This was not the result of any one intervention, but a coordinated programme of taxes, bans in the workplace and public spaces, and advertising restrictions. Gradually, the 'normal' behaviour changed. Someone wishing to smoke today would probably ask their companions for permission; a generation ago, they would have offered them a cigarette.

When it comes to changing the culture of science, the UK Reproducibility Network is not alone. Numerous initiatives now link members of the research community to support robust, transparent research. Examples include the Center for Open Science in the United States (founded in 2013), the QUEST Center in Germany (founded in 2017), the Research on Research Institute with eight participating countries, launched this year, and a proliferation of grass-roots networks of researchers in many countries.

But these cultural changes might falter. Culture eats strategy for breakfast – grand plans founder on the rocks of implicit values, beliefs and ways of working. Top-down initiatives from funders and publishers will fizzle out if they are not implemented by researchers, who review papers and grant proposals. Grass-roots efforts will flourish only if institutions recognize and reward researchers' efforts.

Funders, publishers and bottom-up networks of researchers have all made strides. Institutions are, in many ways, the final piece of the jigsaw. Universities are already investing in cutting-edge technology and embarking on ambitious infrastructure programmes. Cultural change is just as essential to long-term success.

News in brief

SCIENCE-FUNDING PLEDGES AHEAD OF 'BREXIT ELECTION'

As *Nature* went to press, the United Kingdom's three main political parties had each pledged to increase spending on science ahead of the 12 December general election, which was largely overshadowed by the country's 2016 vote to leave the European Union.

Conservative Prime Minister Boris Johnson called the election after Parliament was unable to agree on how to move forward in the wake of the Brexit vote.

During the campaign, the three main parties diverged widely on their plans for the impending divorce, with the Conservatives vowing to take the country out of the bloc by 31 January, Labour promising a second 'people's vote' with an option to remain in the EU and the Liberal Democrats vowing to stop Brexit altogether.

The spectre of Brexit "looms large over science", says James Wilsdon, director of the Research on Research Institute at the University of Sheffield, UK. As well as its wide political and economic effects,

Brexit's changes to freedom of movement and participation in EU programmes could have major impacts on scientists.

But although Brexit was the dominant campaign issue, the Conservatives, Labour and the Liberal Democrats had all promised in their manifestos to invest more money in research.

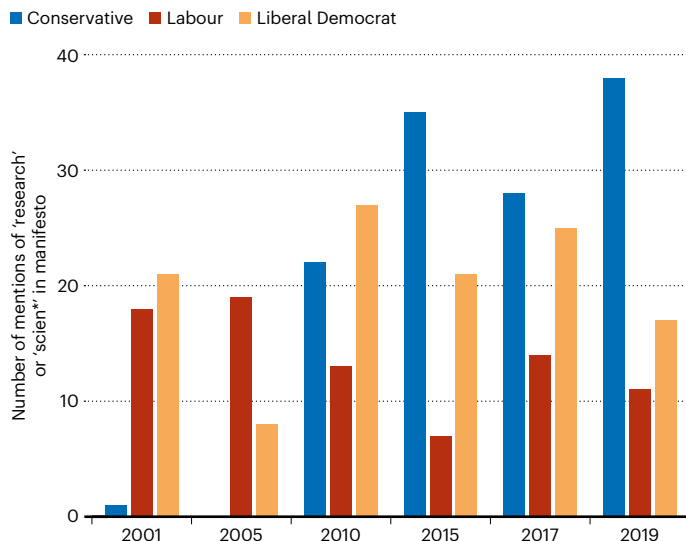
The United Kingdom currently spends about 1.7% of its gross domestic product (GDP) on research and development, totalling about £7.5 billion (US\$9.8 billion) per year.

The Conservative Party – whose manifesto mentioned the words 'research' and 'science' more times than ever before, and more than those of the other two main parties (see 'Political push') – was pressing for the lowest increase. It pledged to spend 2.4% of GDP on science.

The other two parties were targeting 3%, with Labour pledging to hit the goal by 2030 and "create an innovation nation". The Liberal Democrats promised an "interim target" of 2.4% by 2027.

POLITICAL PUSH

Mentions of science in the Conservative Party manifesto are at their highest in two decades.



INDONESIA FIRE ANALYSIS GETS HEATED

How much did Indonesia burn this year? An international research organization has taken down an online report that suggested fires burnt more than 1.6 million hectares of land in the country in 2019, 40% more than the government calculated for the same period.

The Indonesian government had criticized the analysis, by the Center for International Forestry Research (CIFOR), saying that it relied on satellite data that hadn't been confirmed with ground observations.

Raffles Panjaitan, director of forest and land-fire management at the Ministry of Environment and Forestry in Jakarta, says the ministry used satellite data and ground observations to calculate that just under 950,000 hectares were burnt between January and October. The ministry relies on imagery from NASA's Landsat 8 Earth-observation satellite.

CIFOR, based in Bogor, Indonesia, used time-series imagery from the European Space Agency's Sentinel-2 Earth-observation satellites. On 6 December, the group removed the analysis from its blog, saying it should have had it peer reviewed before posting.

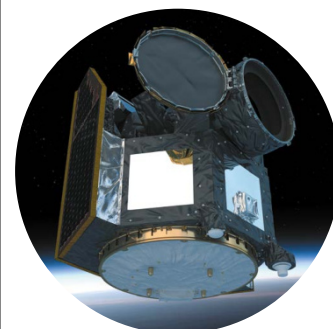
Three days earlier, CIFOR landscape ecologist David Gaveau had told *Nature* that the discrepancies might stem from differences in the satellite imagery the two groups used.

SPACE TELESCOPE HERALDS NEW ERA OF EXOPLANET SCIENCE

The European Space Agency is set to launch an exoplanet telescope that will perform detailed studies of hundreds of known worlds beyond the Solar System. Called CHEOPS (Characterising Exoplanet Satellite), the €50-million (US\$55-million) spacecraft is scheduled for launch from Kourou, French Guiana, on 17 December.

Equipped with a single camera, CHEOPS will peer at stars already known to host exoplanets. By observing the dip in the brightness of a star's light as a planet passes in front of it, astronomers will work out the sizes of these worlds and study some of their atmospheres, providing crucial information on the formation and evolution of a variety of exoplanets. Over its 3.5-year scientific mission, which will begin in April 2020, the telescope will study between 300 and 500 worlds.

The telescope marks a departure from previous exoplanet missions because it is the first designed to study – rather than merely find – alien worlds. "We're moving from discovery to characterization," says Kate Isaak, a project scientist on the mission at the European Space Research and Technology Centre in Noordwijk, the Netherlands.





Cave hunt scene might be world's oldest story

A cave painting in Indonesia portraying a hunt might be the oldest recorded story in human history. The scene – discovered in 2017 and now dated to at least 43,900 years ago – is more than 20,000 years older than depictions of clear narrative scenes in cave art from Europe, such as those found in the Lascaux Cave in France. One panel (pictured) from the Indonesian cave on Sulawesi shows a buffalo called an anoa being hunted by smaller figures that might be human–animal chimaeras. Such figures, also known as therianthropes, point to a human capacity to imagine and depict objects that do not exist in the natural world, says Adam Brumm, an archaeologist at Griffith University in Brisbane, Australia, whose team described the art (M. Aubert *et al.* *Nature* <https://doi.org/10.1038/s41586-019-1806-y>; 2019). Previously, the oldest-known therianthrope was a 40,000-year-old human–lion ivory statue from Germany. The Sulawesi cave and its paintings were found by a local archaeologist and caver named Hamrullah. “We’ve seen hundreds of rock art sites in this region but we’ve never seen anything like a hunting scene,” adds Brumm.



NASA IMAGES REVEAL CRASHED INDIAN MOON LANDER

Images from NASA's Lunar Reconnaissance Orbiter have been used to confirm the final resting place of India's Moon lander, Vikram. The country's space agency lost contact with the craft just moments before it was supposed to touch down on 7 September.

A few days later, the Indian Space Research Organisation said it had located the lander, but did not release any images of the site.

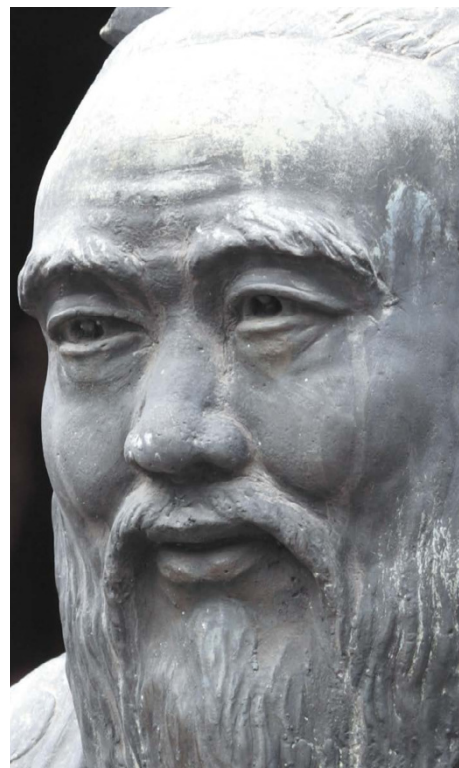
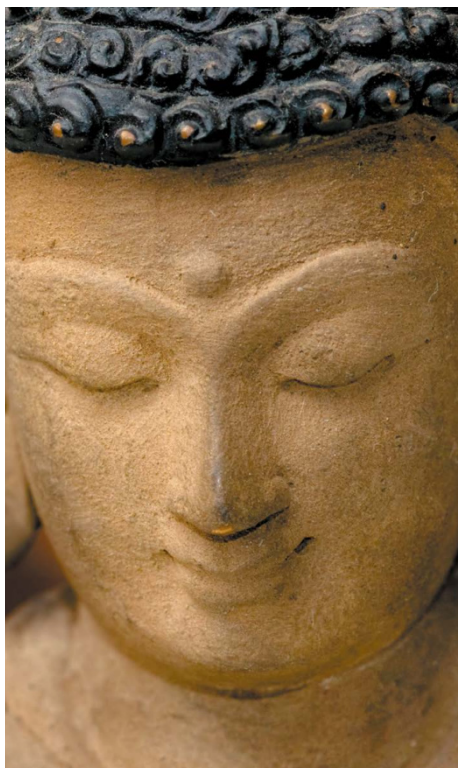
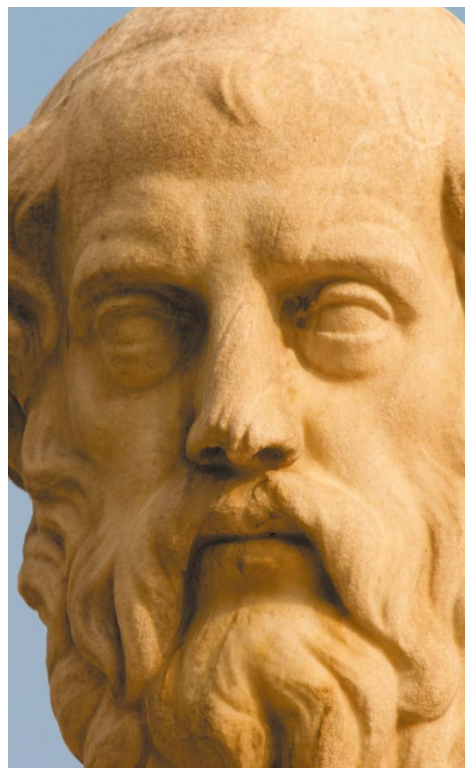
Just over two weeks after the mishap, the team that manages the Lunar Reconnaissance Orbiter's camera posted on its website an image of Vikram's intended landing site near the lunar south pole. It was then contacted by an engineer in Chennai, Shanmuga Subramanian, who had spotted features in the image that looked like debris.

The team was able to confirm that the features were debris from the craft by comparing a more detailed image taken before the landing with photos taken after. The photos reveal a large crater from the main impact and a trail of debris spread across several kilometres.

Vikram was supposed to touch down and explore the area nearby.

The craft was India's first attempt to ‘soft’ land on the Moon, as part of the Chandrayaan-2 mission.

News in focus



Plato (left), Buddha and Confucius (right) were key players in the emergence of an enlightened era, according to philosopher Karl Jaspers.

WHEN DID SOCIETIES BECOME MODERN? 'BIG HISTORY' DASHES IDEA OF AXIAL AGE

Humanity's supposed singular transition to modernity in the first millennium BC was much messier than previously thought, finds sweeping study of historical data.

By Laura Spinney

It's an idea that has been influential for more than 200 years: around the middle of the first millennium BC, humanity passed through a psychological watershed and became modern. This 'Axial Age' transformed an archaic world of divine rulers, slavery and human sacrifice into a more enlightened era that valued social justice, family values and the rule of law. The appeal of the general concept is such that some have claimed humanity is now passing through a second Axial Age – driven by rapid population growth and technological change. Yet according to the largest-ever cross-cultural survey of historical and archaeological data, the first of these ages never happened.

Major changes did take place in how humans understood their place in the Universe and their relationships with each other, finds the analysis. But sometimes these societal shifts happened earlier than the first millennium BC, and sometimes later. And they did not always occur in the societies typically considered 'axial' – what is now Greece, Israel–Palestine, Iran, India and China – although they did take place in some other civilizations. "We couldn't find any consistent Axial Age that was confined to those five societies," says Jenny Reddish, an anthropologist at the Complexity Science Hub in Vienna and one of the survey's authors.

The work, published this week as a 500-page book entitled *Seshat History of the Axial Age*, spotlights the 'big data' approaches to history that have become popular in the past decade.

These can complement the very specialized, detailed work of standard historians with a broad-brush, comparative approach to the evolution of societies that are widely separated in time and space. The current finding is likely to be followed by many other studies that address the origins of complex societies using these new techniques.

Lone innovation?

Although the Axial Age concept dates back to the eighteenth century, its best-known proponent made his case in the 1940s. In 1948, German–Swiss philosopher Karl Jaspers wrote that between 800 and 200 BC, the five aforementioned societies independently embraced moral universalism, or the idea that people are morally bound to each other

News in focus

by their common humanity. This step, Jaspers argued, was prompted by the teachings of lone cultural innovators – Plato, the Hebrew prophets, Zarathustra, Buddha and Confucius – and initiated a trend that eventually touched the entire globe.

The latest work challenges this idea. It builds on a survey published in 2018 and led by anthropologist Daniel Mullins at the University of Oxford, UK, and both draw on a pioneering historical and archaeological database called Seshat – after the ancient Egyptian goddess of recordkeeping. The 2018 study reported that the picture of a simultaneous axial shift was much less clear than scholars had suggested (D. A. Mullins *et al. Am. Sociol. Rev.* **83**, 596–626; 2018). Its sample comprised the five societies mentioned by Jaspers and five others the authors had selected, which thrived between 3,000 BC and AD 2,000 – in modern-day Italy, Turkey, Egypt, Cambodia and Japan.

The latest survey – conducted by some of the same scholars – expands the sample to include societies in southeast Asia, Polynesia, West Africa and North and South America during the same period. The work, which has not been peer reviewed, comes to the same conclusion as the 2018 analysis: it shows “beyond reasonable doubt, that the pattern we need to explain is much messier than Jaspers’ model of a moral and intellectual revolution in first-millennium BCE Eurasia”, says Ian Morris, a historian and archaeologist at Stanford University in California, who was not involved in the survey but wrote the book’s foreword.

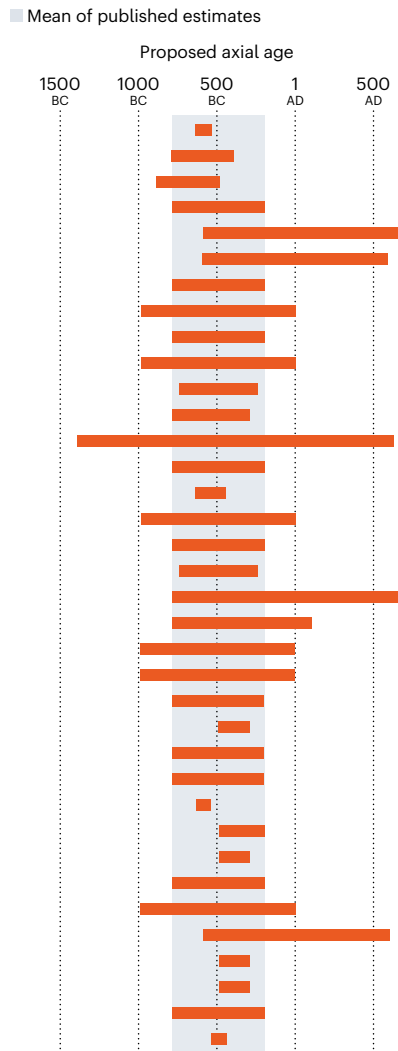
Digging for data

Many historical databases are now under development, but Seshat is among the oldest, the most ambitious and the fastest growing. Founded in 2011 by historians, anthropologists and mathematicians and funded by UK and European Union research grants, among other sources, it now stores information on more than 450 societies going back as far as 4000 BC. Research assistants input data from primary and secondary historical sources using a formula that allows for comparison across time and space while factoring in disagreement or uncertainty. Specialists – such as historians and archaeologists – then validate the work. For the latest survey, the researchers drew up a list of 12 proxy measures that scholars widely agree on for components of ‘axial transformation’ in a society. These include the presence of a formal legal code, belief in an all-knowing supernatural being and the existence of full-time bureaucrats who held rulers accountable. The researchers then tracked the proxies through time and across the expanded list of societies.

Their findings throw up a wide spectrum of times and places for the emergence of axial features. There has long been debate as to whether Egypt – a precociously sophisticated society – should have been included

ANCIENT TO MODERN

Historians have long suggested that an ‘Axial Age’ – a supposed simultaneous shift to modernity in several ancient societies – occurred some time in the first millennium BC, but their estimates of when this happened vary widely. Now, a team of scholars suggests that there was no such neat transition.



Estimates published from 1873 to 2016.

“The disagreement is less about the timing, the place, than about the causal mechanism.”

in Jaspers’ original list, but the survey found clear evidence of axiality there, long before Jaspers’ ‘age’. “Starting around 1200 BCE, in the Ramesside period, you see lots of changes in religious practice,” says Joe Manning, a historian at Yale University in New Haven, Connecticut, who co-authored the Egypt chapter of the study. “It has been called the age of personal piety.” In Anatolia, meanwhile, the Hittites were applying the rule of law universally in the second millennium BC. By contrast, China, although included in the original group of axial civilizations, didn’t pivot until several

centuries after 200 BC, the latest findings suggest.

According to the Seshat team, the data also clearly undermine another of Jaspers’ key claims: that innovation arose independently in the five core societies, which he referred to as “islands of light”. These societies were engaged in a “ton of cross-cultural exchange”, says historian and Seshat project manager Daniel Hoyer at George Brown College in Toronto, Canada. “The Rabbinical tradition and even Plato’s writings aren’t really conceivable without the Zoroastrianism and Egyptian moral ideals and Hittite legalism that went before.”

Differing definitions

The study has its critics. Hans Joas, a sociologist of religion at the Humboldt University of Berlin, says that the authors distort the existing literature to make their point. “It’s unfortunate that people still use the term ‘Axial Age’ as if it mainly referred to a mysterious simultaneity,” says Joas. “That claim has long been given up.”

Existing definitions of the Axial Age differ widely (see ‘Ancient to modern’), however, and the current survey strengthens others’ suspicion that the concept has served its purpose. “It was useful initially for signalling that antiquity was not one homogeneous period, that there was important evolution between the archaic world and late antiquity,” says Nicolas Baumard, who studies cultural evolution at the Ecole Normale Supérieure in Paris. But according to him, the debate has moved on – and the Seshat survey vindicates that. Now, says Baumard, “the disagreement is less about the timing, the place, than about the causal mechanism”.

The survey is unlikely to be the last word on the Axial Age, if only because psychological states are hard to quantify. There is also an ongoing and at times vitriolic debate about the best way to build historical databases. For example, a rival to Seshat, the Database of Religious History, gets specialists such as historians to input the data – rather just oversee non-specialists’ efforts.

Such rifts notwithstanding, says Morris, big-data approaches open up the exciting possibility of testing different theories about what caused axial-type changes. Were they, as he, Baumard and others have proposed, a consequence of increasing affluence? Or were they, as the Seshat group contends in its latest work, a way of maintaining social cohesion as societies became more complex?

The books’ claims might have implications for suggestions that humanity is in, or approaching, a new axial age. Morris says such claims aren’t surprising given the speed with which technology is transforming societies, but the nature of any new axial age is every bit as fuzzy as that of the first – as revealed by the Seshat survey.

SOURCE: D. A. MULLINS *ET AL. AM. SOCIOLOGICAL REVIEW* **83**, 596–626 (2018).



The earthquake that devastated San Francisco, California, in 1906 arose from the San Andreas fault.

TWO BIG QUAKE FAULTS MIGHT BE LINKED

Analysis suggests that quakes on the Cascadia fault off California can trigger shaking on the San Andreas.

By Alexandra Witze

Two of North America's most fearsome earthquake zones could be linked.

A controversial study argues that at least 8 times in the past 3,000 years, quakes made a one-two punch off the west coast of the United States. A quake hit the Cascadia fault off the coast of northern California, triggering a second quake on the San Andreas fault just to the south. In some cases, the delay between the quakes might have been decades long.

The study suggests that Cascadia, which scientists think is capable of unleashing a

magnitude-9 earthquake at any time, could set off quakes on the northern San Andreas, which runs under the San Francisco Bay Area. Several earthquake scientists told *Nature* that more work is needed to confirm the provocative idea. Researchers have long considered the two faults seismically separate.

Chris Goldfinger, a geologist and palaeoseismologist at Oregon State University in Corvallis, will present the findings on 13 December at a meeting of the American Geophysical Union in San Francisco. "This is mostly a circumstantial case," he says. "I don't have a smoking gun."

Goldfinger and his colleagues first suggested in 2008 that earthquakes in the

southern part of Cascadia could trigger quakes on the northern San Andreas (C. Goldfinger *et al. Bull. Seismol. Soc. Am.* **98**, 861–889; 2008). The scientists reported finding layers of churned-up, sandy sediment in sea-floor cores drilled offshore. These layers, called turbidites, usually form when earthquakes cause underwater landslides. The researchers found turbidites in Cascadia that seemed to form just before similar turbidites near the San Andreas – perhaps as a Cascadia quake triggered a San Andreas one.

But it was hard to pinpoint exactly when the turbidites had formed, and Goldfinger knew he needed more evidence. Now he has data from seven cores drilled offshore in southern Cascadia and seven cores drilled near the northern San Andreas. The two sites are around 100 kilometres apart – close enough to feel shaking from both faults.

At eight places in both sets of cores, Goldfinger spotted unusual, two-layered turbidites and realized that they were telling him something new. The two-layered turbidite "has to be two quakes recorded together", he says. As Goldfinger sees it, a Cascadia quake shook the coastline first, causing landslides that show up

News in focus

in both sets of cores as the first layer of turbidites. Then, at some later point, the northern San Andreas also shook, causing the second turbidite layer to form.

“This story is pretty convincing,” says Jason Patton, an engineering geologist with the California Geological Survey in Sacramento who was a co-author on the 2008 paper. “Cascadia turbidites are covered by San Andreas turbidites, so the Cascadia turbidites were deposited first.”

Others are reserving judgement. Turbidites show that the ground shook at some point in the past, but it’s difficult to tell exactly when or where those quakes happened, says Joan Gombert, a seismologist at the US

Geological Survey in Seattle, Washington. “All this uncertainty leaves multiple, equally plausible interpretations on the table – most of which are not sensational,” she says.

Ross Stein, a seismologist with the earthquake-preparedness firm Temblor in Redwood City, California, wants to see detailed modelling of how stress from the Cascadia fault might be transferred to the northern San Andreas. Scientists generally agree that a large earthquake can sometimes trigger another on a nearby fault. But it’s not clear whether that might happen between southern Cascadia and the northern San Andreas, Stein says.

This week at the conference, Goldfinger says, “I’m just going to lay out the case.”

detentions and other human-rights violations in the province. The Chinese government says it is conducting a re-education campaign to quell what it calls a terrorist movement.

“We are very concerned about research which involves consent from vulnerable populations,” says a spokesperson from Springer Nature (*Nature’s* news team is editorially independent of its publisher).

The publishers’ announcements, which *The New York Times* reported on 4 December, follow rising concerns about the publication of such work. Last week, Yves Moreau, a computational biologist at the Catholic University of Leuven in Belgium, wrote an opinion article in *Nature* warning of the dangers that accompany the proliferation of DNA profiling and calling for all unethical work in biometric research to be retracted.

Springer Nature said that it would add notes of concern about consent to two papers^{1,2} that reported studies using DNA from hundreds of Uyghurs to predict height or facial shape. One, published in *Human Genetics*², was highlighted in a separate *New York Times* article that described worries that the participants hadn’t given informed consent.

Both papers state that volunteers gave consent, and that the studies were approved by an ethics committee from the Institute of Forensic Science, which is affiliated with China’s police and security authority.

“We are ordinary forensic scientists who carry out forensic research following the scientific research ethics norms,” said Caixia Li of the Institute of Forensic Science in Beijing, a co-author of both papers^{2,3}, in an e-mail to *Nature’s* news team. He said that “all individuals provided written informed consent”.

Moreau says that it’s hard to see how Uyghur peoples could give free, informed consent to DNA or facial-recognition work – given that so many people in that ethnic group have been sent to internment camps (which China calls education facilities).

Springer Nature has identified a number of other ‘papers of concern’ published by its journals, the spokesperson adds, which are being investigated. And it has updated its guidance about the need to gain explicit and informed consent in studies that involve clinical, biomedical or biometric data from people.

Moreau says: “Expressions of concerns are a welcome first step, but this is only meaningful if it is the start of a large-scale ethical review of all forensic population-genetic research on Chinese populations and of all biometric research.”

Wiley, meanwhile, said it was opening a formal investigation into an article that described an analysis of a database of photos of Uyghur, Tibetan and Korean people using various facial-recognition algorithms⁴. In September, four researchers, including Moreau and Jack Poulson of the advocacy group Tech Inquiry

PUBLISHERS REVIEW RESEARCH ON CHINESE MINORITY GROUPS

Springer Nature and Wiley have concerns about the ethics of papers on genetics and facial recognition.

By Richard Van Noorden & Davide Castelvecchi

Two science publishers are reviewing the ethics of research papers in which scientists backed by China’s government used DNA or facial-recognition technology to study minority groups in the country, such as the predominantly

Muslim Uyghur population.

Springer Nature (which publishes *Nature*) and Wiley want to check that the study participants gave informed consent, after researchers and journalists raised concerns that the papers were connected to China’s heavy surveillance operations in the northwestern province of Xinjiang. China has attracted international condemnation – and US sanctions – for mass



Officers patrol in China’s Xinjiang region, where there have been mass detentions.

in Toronto, Canada, had asked Wiley to retract the paper because of the potential for abuse of facial-recognition technology, and the “racial overtones of the authors’ language”. One of the paper’s co-authors is affiliated with Curtin University in Perth, Australia, which this month also requested that Wiley review the paper.

A spokesperson for Wiley said that the publisher was now re-evaluating the paper, following an initial review that found the journal had followed the existing guidelines.

In their September letter, the researchers noted that, according to guidelines laid down by the Committee on Publication Ethics, a London-based publisher-advisory body, papers can be rejected on ethical grounds even if they come with approvals from an ethics committee (as the Wiley paper did). The guidelines also say that journals should take special care when the research is conducted on “vulnerable groups”, which Springer Nature has emphasized in its latest editorial policies.

Widespread concerns

There are numerous papers that report the use of biometric technology to study Uyghur and other minority ethnic groups in China. Moreau wrote in his opinion article that he had identified 40 articles co-authored by members of the Chinese police in 3 leading forensic-genetics journals – published by Springer Nature and by Elsevier – that describe the DNA profiling of Tibetans and people from Muslim minorities. A spokesperson for Elsevier said that the company is producing more comprehensive guidelines for the publication of genetic data, but that it was “unable to control the potential misuse of population data articles” by third parties after publication.

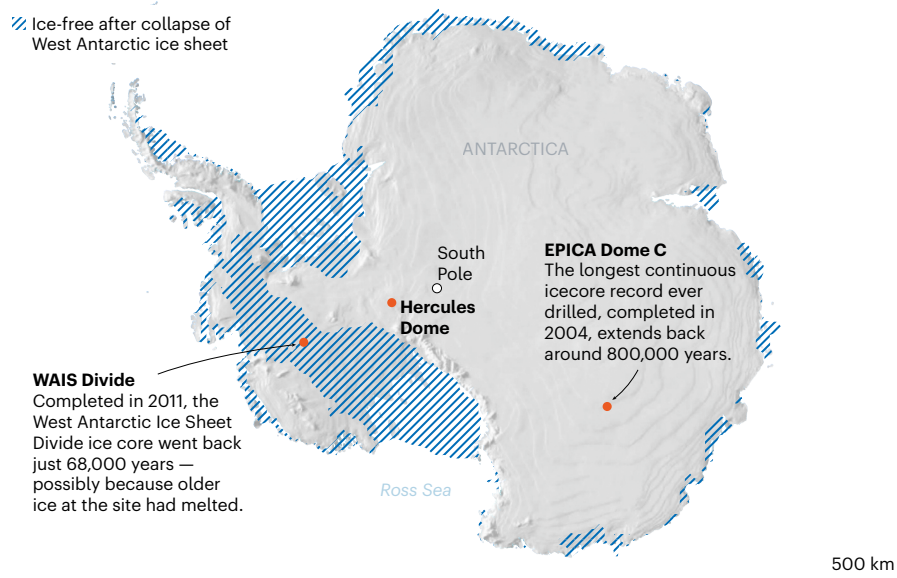
Springer Nature also said that it would tighten its oversight of the academic-conference proceedings that it publishes. Journalists have previously raised concerns about numerous conference papers that describe studying Uyghur groups, including manuscripts from a biometrics conference held in Xinjiang in 2018 that Springer Nature published as a book⁵.

The Institute of Electrical and Electronics Engineers (IEEE) has also published conference proceedings describing facial-recognition analyses in Uyghur populations (see, for example, refs 6 and 7). “IEEE is committed to reviewing our policies to ensure more consistent application of this process across the full range of IEEE publications,” IEEE president and chief executive José Moura wrote in an e-mail.

1. Moreau, Y. *Nature* **576**, 36–38 (2019)
2. Li, Y. et al. *Hum. Genet.* **138**, 681–689 (2019).
3. Jing, X. et al. *Int. J. Legal Med.* **133**, 1667–1679 (2019).
4. Wang, C., Zhang, Q., Liu, W., Liu, Y. & Miao, L. *WIREs Data Mining Knowl. Discov.* **9**, e1278 (2019).
5. Zhou, J. et al. (eds) *Biometric Recognition. CCBP 2018. Lecture Notes in Computer Science* (Springer, 2018).
6. Zuo, H., Wang, L. & Qin, J. in 2017 *Int. Conf. Machine Vis. Inf. Technol. (CMVIT)* 7–11 (2017).
7. Duan, X. et al. in 2010 *2nd Int. Conf. Adv. Computer Control* 125–130 (2010).

CLIMATE CLUES

The ice at Hercules Dome could help scientists confirm whether the West Antarctic ice sheet collapsed 130,000 years ago. That could help them assess the threat of a similar collapse in the coming centuries.



THE HUNT FOR ICE THAT WITNESSED WEST ANTARCTICA'S COLLAPSE

Ice at Hercules Dome site could reveal how susceptible the region is to warming.

By Jeff Tollefson

Sometime this month, scientists in Antarctica plan to start up their snowmobiles and begin radar surveys of a thick ridge of ice called Hercules Dome. The dome – which sits 400 kilometres from the South Pole, between East and West Antarctica – could provide crucial clues to the future of the continent’s vast ice sheet.

The surveys are intended to guide the drilling of the United States’ next deep ice core. Glaciologists hope to retrieve a detailed climate record of a period 116,000 to 130,000 years ago, when temperatures as little as 1°C warmer than today’s are thought to have driven the collapse of the ice covering West Antarctica.

A better understanding of what happened then could help scientists to predict the behaviour of West Antarctica as climate change intensifies. The pace at which the region’s glaciers are flowing to the sea has increased in recent years, and many scientists fear that rising temperatures have triggered runaway melting. West Antarctica’s ice contains enough water to raise sea levels by more than 3 metres.

“If the West Antarctic ice sheet collapsed, Hercules Dome would basically be waterfront property,” says Eric Steig, a glaciologist at the University of Washington in Seattle who is leading the project. He is hopeful that scientists will be able to see the signal of the region’s long-ago collapse in ice from the dome.

The US National Science Foundation (NSF) has given Steig and his colleagues nearly

“There may well be ice that is two million years old at the bottom. I wouldn’t be surprised.”

US\$630,000 to conduct radar surveys of ice depth and structure across Hercules Dome. That work started in January. The \$1.5-million drilling project could begin as soon as 2022, pending the agency’s approval.

“The Hercules structure will have served as witness to what the atmosphere and what the oceans were doing when the West Antarctic ice sheet collapsed,” says Mary Albert, a glaciologist at Dartmouth College in Hanover, New Hampshire, and head of the US Ice Drilling

News in focus

Program, which advises the NSF.

Bad weather and unexpectedly rough terrain made it difficult to land the NSF's Hercules cargo plane at Hercules Dome last year, limiting scientists' ability to conduct radar surveys. Nonetheless, the University of Washington team managed to identify two potential drilling sites where the depth of the ice ranges from 1.6 to 2.8 kilometres, says glaciologist Knut Christianson, who is leading the radar surveys.

This year, he and his colleagues plan to spend around 30 days collecting data – helped along by a newly groomed ice runway. Christianson's team is looking for ice that is solidly frozen all the way to Antarctic bedrock. That increases the odds that the researchers can recover air bubbles trapped in the oldest layers of snow and ice, which accumulated while Earth was cycling through ice ages.

These air bubbles can reveal how the levels of greenhouse gases, trace gases and aerosols in the atmosphere changed from decade to decade, helping researchers to reconstruct past climate. Data from ancient ice at Hercules Dome could help to reveal whether the West Antarctic ice sheet was intact 115,000 to 130,000 years ago, during a brief interlude between ice ages.

There is evidence that sea levels during that period were up to 9 metres higher than today's – a phenomenon that Steig says is hard



A view of Antarctica's vast ice sheet.

to explain without the loss of West Antarctica's ice. But hard evidence for the ice sheet's collapse has been tough to come by.

Cold case

Glaciologists had hoped to capture that ancient warm period in an ice core that they finished drilling in West Antarctica in 2011 (see 'Climate clues'). But the core they extracted

dated back only 68,000 years, possibly because the older ice had melted away. Models suggest there is little or no melting in the deepest ice layers at Hercules Dome, which are likely to be at least 150,000 years old.

"There may well be ice that is two million years old at the bottom," Steig says. "I wouldn't be surprised." But it's not clear how much information scientists can extract from such ancient, and often very compressed, ice.

One of the biggest questions at Hercules Dome is how the weather there changed when the climate warmed 130,000 years ago, says Jeffrey Severinghaus, a palaeoclimatologist at the Scripps Institution of Oceanography in La Jolla, California. Climate models suggest that the collapse of West Antarctica's ice changed air circulation across the region, creating stormier conditions at Hercules Dome.

Severinghaus's research has shown that the prevalence of isotopes of krypton and other trace gases deposited in ice varies with atmospheric pressure. This suggests that scientists should be able to see evidence of storms in ice from Hercules Dome.

Drilling there could provide an unprecedented opportunity to determine what happened during a crucial period of Antarctica's history, Severinghaus says. "This is a good analogue," he adds, "for where humanity is headed right now."

MARIO TAMAI/GETTY



**The week's best science,
from the world's leading
science journal.**

NATURE.COM/NATURE/PODCAST

nature

THE CAUSATION DETECTOR

A technique called Mendelian randomization has become the go-to for drawing lessons from epidemiological data. But are scientists overdoing it? **By David Adam**

In 1812, the British ophthalmologist James Ware relayed a curious finding to the members of the Royal Society in London. Of thousands of young men recruited to regiments of the British army, only six had been turned away for poor vision in 20 years. But up to one-quarter of students about the same age going to the University of Oxford, UK, relied on a hand glass or spectacles¹.

Ware didn't draw any conclusions about cause and effect: that poring over books might contribute to poor eyesight, for example, or that the bespectacled are naturally drawn to academic pursuits. And just as well.

Epidemiologists have long been frustrated by observations that link environmental exposures and health. Myopia is a classic example. Decades of studies show that children who spend the most time at school have the worst eyesight. But the data don't reveal whether schooling makes children myopic or whether myopic kids spend more time at school. Or whether something else, such as socio-economic status, drives both.

Fed up with this logical cul-de-sac, by the turn of this century some epidemiologists had begun suggesting that their field should call it a day. Advances in genetics, they said, could do a better job.





They were half-right. Two decades on, genetics has transformed how people untangle correlation from causation. But it has come to raise epidemiology, not bury it. Genetic differences, it turns out, can help remove confounding variables from analyses, by standing in as proxies for environmental exposure. The technique is called Mendelian randomization.

Scientists have used it to re-evaluate observational data and draw fresh, firmer conclusions on long-standing questions of cause and effect. The analyses have affirmed that low cholesterol levels do not cause cancer², for example, that drinking small amounts of alcohol does not protect the heart³ and that – yes – schooling can make children short-sighted⁴.

“Mendelian randomization in principle is a really, really cool idea. It attempts to solve one of the most daunting challenges in epidemiology,” says Philipp Koellinger, a social-science geneticist at the Free University of Amsterdam.

Gathering momentum

George Davey Smith, a clinical epidemiologist at the University of Bristol, UK, who helped to pioneer the technique, says: “It came about because we were getting desperate and looking for ways of getting better causal inference in epidemiology.” But, he says, there is a downside, too. “The issue is that it became very simple to do.”

He has been urging colleagues not to get carried away with Mendelian randomization. It’s a powerful tool, but one that must be used properly. As genetic data have piled up, a flurry of Mendelian randomization studies have emerged that don’t make the grade. Some have relied on misleading data, and others have failed to sufficiently test the assumptions on which Mendelian randomization relies. It’s time, many in the field say, to tighten things up.

Davey Smith was one of the scientists who suggested that epidemiology might have run its course. Writing in an editorial in the *International Journal of Epidemiology*, he and a co-author pointed out that the observational data on the possible harm or benefits of environmental exposures would repeatedly fail when interventions were tested in randomized controlled trials⁵.

A few years after the article was published, that point came through loud and clear in the high-profile failure of a US\$100-million trial



**THE REPUTATION
OF EPIDEMIOLOGY
WAS COMING UNDER
SCRUTINY.”**

called SELECT, which found that eating selenium supplements did not protect against prostate cancer – despite mountains of epidemiological evidence suggesting that it would⁶.

“It was all rather depressing, and the reputation of epidemiology was coming under scrutiny,” Davey Smith says. Researchers had suggested⁷ as early as 1986 that genetics could improve the interpretations. But it took the growth of genome-wide association studies (GWAS), which link genetic variants to specific traits, for the approach to gain traction. Last year, Davey Smith turned to Mendelian randomization to revisit the selenium–prostate cancer connection.

Using genotype data for tens of thousands of men, the researchers found almost a dozen genetic variants that were associated with naturally higher levels of selenium in the blood⁸. From birth, these people had lived as if they were taking selenium supplements. The scientists could then compare the incidence of prostate cancer in people with these variants to that in a control group without them. That allowed the researchers to focus more squarely on selenium levels and to ward off the influence of lifestyle factors, such as a healthy diet, that might influence both selenium levels and cancer risk. And, because the tendency to have high or low selenium levels was fixed in DNA, the analysis was less troubled by the likelihood of reverse causation: the possibility that early stages of prostate cancer might influence selenium levels.

The analysis found no benefit from selenium⁸, just as the SELECT trial had done⁶.

Data bounty

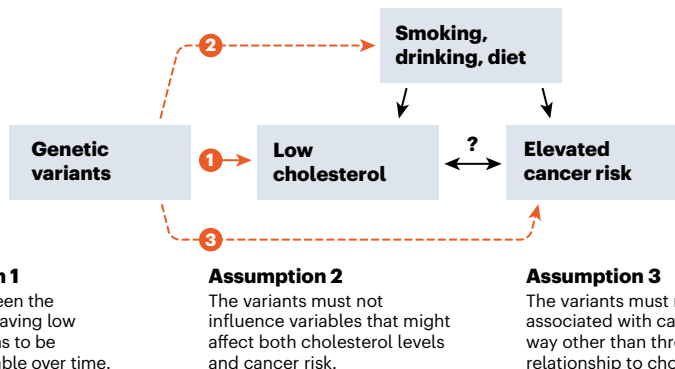
Such results can feed into decisions about whether to launch full clinical trials, Davey Smith argues. And Mendelian randomization can test hypotheses for which it would be unethical or impractical to carry out a trial.

In principle, a Mendelian randomization analysis can be done wherever a genetic variant can be found to naturally mimic the effects of an environmental exposure. And more of those are found every year – especially as millions of people around the world sign up to have their genomes analysed and their health tracked. That gives geneticists the statistical power to identify genetic associations with everything from alcohol consumption to cholesterol levels.

Now, epidemiologists and others are feeding these findings into more Mendelian randomization tests. Data from Scopus and the Web of Science list fewer than 100 papers published per year on the topic in 2010, growing to about 200 by 2015. In 2019, so far, more than 500 papers have used or discussed the method. Researchers have used the tests to tackle a number of questions typically confounded by life’s many variables. Studies have helped to show more definitively that drinking alcohol

GENES AS PROXY

Mendelian randomization uses gene variants to interrogate cause and effect in epidemiological data. For example, researchers can ask whether having low cholesterol increases cancer risk – as some data suggest – by looking at people genetically predisposed to low cholesterol. This strategy can help rule out reverse causation (that cancer lowers cholesterol), and it can bypass some variables that might influence both cancer risk and cholesterol. However, it relies on several assumptions, which must be tested.



can increase the risk of cancer⁹. Meanwhile, having low cholesterol does not², despite some observations to the contrary (see ‘Genes as proxy’).

As a prime example of how Mendelian randomization can help, many researchers point to myopia, which is a rapidly growing public-health issue. It’s been impossible to test its connection to schooling with a randomized

data, they found that being genetically prone to myopia made no difference to how many years people had spent at school. Those who carried the genes associated with educational attainment, however, were significantly more likely to be short-sighted⁴.

Whether through time spent reading, lower levels of natural light or some other factor, time spent in school clearly influences vision, says Denize Atan, an ophthalmologist at Bristol who led the project. The link is so strong, she says, that policymakers and schools should do more to address it.

The open secret

The problem is, critics argue, that not all Mendelian randomization studies are as sound. “You need to have a robust hypothesis and some supporting evidence before you start,” Atan says. A growing number, she says, do not. “You think, ‘Where did they get that idea from?’ It just seems to come out of the blue.” This is a big problem, she adds, because Mendelian randomization allows researchers to seek, find and publish relationships between unfamiliar data sets without any specialist knowledge of the relevant field.

As Sonja Swanson, an epidemiologist at the Erasmus University Medical Center in Rotterdam, the Netherlands, puts it: “It doesn’t take much to just hit the buttons and say ‘here’s a numeric answer to my question’”.

Several epidemiologists say that it’s an open secret in the field that many published Mendelian randomization studies are problematic. “You can get papers published very easily,” says Davey Smith. “Some of the very poor papers are from researchers who don’t understand epidemiological principles.” In 2016, a Mendelian randomization study claimed to have found that high blood levels of C-reactive protein, a liver enzyme associated with inflammation, caused schizophrenia. It suggested that drugs capable of lowering levels of the enzyme in the blood might help to treat people with the disease¹². Davey Smith’s group and

IT DOESN'T TAKE MUCH TO JUST HIT THE BUTTONS AND SAY 'HERE'S A NUMERIC ANSWER'.

controlled trial, because it’s unethical to deliberately keep some children out of school.

A chance to use Mendelian randomization came in 2016, after geneticists published data from two separate GWAS: one looking for genetic signatures related to educational attainment¹⁰; the other looking for genes associated with myopia¹¹. The studies looked at hundreds of thousands of people and found dozens of genetic variants robustly associated with myopia and years spent in school.

The next year, epidemiologists used these variants to explore one of the biggest population data sets around – 488,000 middle-aged and older people who had signed up to the UK Biobank project. Volunteers have their genomes analysed and answer questions on dozens of personal details, including their education and eyesight. When the researchers – at the University of Cardiff and the University of Bristol in the United Kingdom – used Mendelian randomization to analyse the

another did similar analyses, and found that, in fact, C-reactive protein had a protective effect against schizophrenia¹³. Davey Smith and his co-authors suggested that there were issues with how the initial group had combined genetic data sets, and the 2016 paper was eventually retracted.

Epidemiologists have also criticized¹⁴ Mendelian randomization analyses that claimed to have found that smoking while pregnant causes dramatic declines in baby birthweights¹⁵ and substantially increases the risk of orofacial clefts in offspring¹⁶.

The problem, they say, is that the genetic variants used as proxies for smoking behaviour were identified in what are known as candidate-gene studies, in which researchers evaluate a few genes that they suspect are involved in a behaviour such as smoking. Results from such studies can be unreliable because they are biased towards finding some effects in the genes being examined. The variants that the authors used in their Mendelian randomization haven't shown up in larger, more comprehensive GWAS.

George Wehby, a health-policy researcher at the University of Iowa in Iowa City, who led the smoking projects, says that the work was done before better data were available. "I agree that these wouldn't be the first choice," he says, "given current knowledge about genetics of smoking from large GWAS."

Against common sense

To an economist, Mendelian randomization looks a lot like something called instrumental variable analysis, in which a variable referred to as the instrument is used to help unpick hidden relationships between two other observations. "When we saw that epidemiologists were using genes as instrumental variables, we were both intrigued and said, 'Wait a moment!'," says Koellinger. Such analyses are built on assumptions that need to be carefully scrutinized.

One central assumption in Mendelian randomization is that the genetic variants must not affect the outcome in any other way. For example, there is a variant of the gene that encodes the enzyme aldehyde dehydrogenase (ALDH2) that disrupts metabolism of alcohol. When people with this variant drink, they tend to feel nauseous, and so it's associated with lower levels of alcohol consumption. That might seem a plausible way to test, for example, whether drinking raises blood pressure, because those who carry the variant generally drink less than those who don't.

The problem is that ALDH2 also influences how likely someone is to smoke¹⁷, which independently influences blood pressure. This phenomenon, known as genetic pleiotropy, can invalidate Mendelian randomization results. And that creates a problem, because the extent of pleiotropy isn't fully realized for

many genes.

Another assumption is that a given genetic variant has a strong effect. As bigger and more powerful GWAS dredge up weaker genetic links to different traits, this assumption becomes harder to test.

A 2015 review by epidemiologists in the Netherlands of 178 published Mendelian randomization studies found that fewer than half adequately discussed these assumptions¹⁸. "As these assumptions are crucial for the validity of Mendelian randomization studies, they should always be discussed in the specific context of the study," the researchers argue¹⁸.

Mendelian randomization is also subject to a distinct source of bias – one that's a matter of life and death. People can die only once. This issue would complicate, for example, an analysis of deaths from stroke. Such deaths tend to occur in older people, so a study of strokes will typically recruit people who have already survived conditions that affect younger people, such as heart disease. Because stroke and heart disease have common causes, such as high cholesterol (and therefore common therapies, including statins), this survivor bias can throw up some misleading results.

To demonstrate the effects of this bias, Mary Schooling, a public-health epidemiologist at the City University of New York, ran Mendelian randomization tests in which gene variants linked to reduced cholesterol stood in for statin use. People with this beneficial inheritance have fewer heart attacks early in life and live to an age at which the risk of stroke rises. So, the study concluded that cholesterol-lowering statins would actually cause strokes¹⁹.

"It didn't make any sense," Schooling says. Proper randomized controlled trials aren't confused in this way: they show that statins protect against stroke. But Mendelian randomization shows a survivor bias that must be identified and corrected for.

Battling biases

"Every single method can be biased," says Davey Smith. Mendelian randomization, he says, is not intended to replace randomized controlled trials, but, alongside other sources, including observational studies, they can add to the evidence available to help make an informed decision. Now, researchers are looking for ways to improve them.

One way is to identify and correct for some of the biases, and to apply statistical tools for testing the strength of the assumptions. Davey Smith points to papers that can help researchers to assess the quality of Mendelian randomization studies for themselves²⁰.

Better organization of data can help, too. Unbiased analyses assume that genes are randomly distributed, but some genes are known to cluster in geographical regions²¹. Already, genotype data sets are becoming available that are grouped by extended families, and

Mendelian randomization studies of these data are identifying, for example, that height and body mass index might not influence educational attainment as much as previous studies had suggested²².

By comparing the results of within-family and population-based studies, geneticists can help to distinguish the roles of nature and nurture in a given trait. "Particular genes are correlated with particular features of the local environment. And if you want to use genes for causal inference you need to break that link," says Koellinger.

This kind of accuracy is important if researchers want to harness the growing torrent of genetic information for public-health and policy recommendations. But even these tools need to be improved and supplemented.

Ware's observations 200 years ago on the eyesight of students and soldiers have been explained through a genetic lens that no one could have imagined at the time. Ironically, it took the British army another century to accept recruits who need to wear glasses, and to change its standards for what it considers adequate vision. Even during the First World War, some authorities argued it didn't matter if a British soldier couldn't see clearly what he was shooting at, as long as he could "fire in the right direction"²³.

Statistical tools for epidemiology are improving. And although Mendelian randomization does not always offer perfect clarity, it might, at least, point researchers in the right direction.

David Adam is a freelance journalist based near London.

1. Ware, J. *Phil. Trans. R. Soc. London* **103**, 31–50 (1813).
2. Benn, M., Tybjaerg-Hansen, A., Stender, S., Frikke-Schmidt, R. & Nordestgaard, B. G. *J. Natl Cancer Inst.* **103**, 508–519 (2011).
3. Millwood, I. Y. et al. *Lancet* **393**, 1831–1842 (2019).
4. Mountjoy, E. et al. *Br. Med. J.* **361**, k2022 (2018).
5. Davey Smith, G. & Ebrahim, S. *Int. J. Epidemiol.* **30**, 1–11 (2001).
6. Lippman, S. M. et al. *J. Am. Med. Assoc.* **301**, 39–51 (2009).
7. Katan, M. B. *Lancet* **327**, 507–508 (1986).
8. Yarmolinsky, J. et al. *J. Natl Cancer Inst.* **110**, 1035–1038 (2018).
9. Pierce, B. L., Kraft, P. & Zhang, C. *Curr. Epidemiol. Rep.* **5**, 184–196 (2018).
10. Okbay, A. et al. *Nature* **533**, 539–542 (2016).
11. Pickrell, J. K. et al. *Nature Genet.* **48**, 709–717 (2016).
12. Inoshita, M. et al. *Sci. Rep.* **6**, 26105 (2016).
13. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. *Int. J. Epidemiol.* **45**, 1717–1726 (2017).
14. Taylor, A. E. et al. *Econ. Hum. Biol.* **13**, 99–106 (2014).
15. Wehby, G. L. et al. *Biodemography Soc. Biol.* **57**, 3–32 (2011).
16. Wehby, G. L. et al. *Health Serv. Outcomes Res. Methodol.* **11**, 54–78 (2011).
17. Masaoka, H. et al. *Drug Alcohol Depend.* **173**, 85–91 (2017).
18. Boef, A. C. G., Dekkers, O. M. & le Cessie, S. *Int. J. Epidemiol.* **44**, 496–511 (2015).
19. Schooling, C. M. et al. preprint at bioRxiv <https://doi.org/10.1101/716621> (2019).
20. Davies, N. M., Holmes, M. V. & Davey Smith, G. *Br. Med. J.* **362**, K601 (2018).
21. Abdellaoui, A. et al. *Nature Hum. Behav.* <https://doi.org/10.1038/s41562-019-0757-5> (2019).
22. Brumpton, B. et al. Preprint at BioRxiv <https://doi.org/10.1101/602516> (2019).
23. Cubitt, B. B. *Br. J. Ophthalmol.* **2**, 35–40 (1918).



A THEORY IN CRISIS

A promising field of research on social behaviour floundered after investigators couldn't repeat key findings. Now researchers are trying to establish what's worth saving. **By Tom Chivers**

Three years ago, a team of psychologists challenged 180 students with a spatial puzzle. The students could ask for a hint if they got stuck. But before the test, the researchers introduced some subtle interventions to see whether these would have any effect. The psychologists split the volunteers into three groups, each of which had to unscramble some words before doing the puzzle. One group was the control, another sat next to a pile of play money and the third was shown scrambled sentences that contained words relating to money.

The study, published this June¹, was a careful repeat of a widely cited 2006 experiment². The original had found that merely giving students subtle reminders of money made them work harder: in this case, they spent longer on the puzzle before asking for help. That work was one among scores of laboratory studies which argued that tiny subconscious cues can have drastic effects on our behaviour.

Known by the loosely defined terms 'social priming' or 'behavioural priming', these studies include reports that people primed with 'money' are more selfish³; that those primed with words related to professors do better on quizzes³; and even that people exposed to something that literally smells fishy are more likely to be suspicious of others⁴.

The most recent replication effort¹, however, led by psychologist Doug Rohrer at the University of South Florida in Tampa, found that students primed with 'money' behave no differently on the puzzle task from the controls. It is one of dozens of failures to verify earlier social-priming findings. Many researchers say they now see social priming not so much as a way to sway people's unconscious behaviour, but as an object lesson in how shaky statistical methods fooled scientists into publishing irreproducible results.

This is not the only area of research to be dented by science's 'replication crisis'. Failed replication attempts have cast doubt on findings in areas from cancer biology to economics. But so many findings in social priming have been disputed that some say the field is close to being entirely discredited. "I don't know a replicable finding. It's not that there isn't one, but I can't name it," says Brian Nosek, a psychologist at the University of Virginia in Charlottesville, who has led big replication studies. "I've gone from full believer to full sceptic," adds Michael Inzlicht, a psychologist at the University of Toronto, Canada, and an associate editor at the journal *Psychological Science*.

Some psychologists say the pendulum has swung too far against social priming. Among these are veterans of the field who insist that their findings remain valid. Others accept that many of the earlier studies are in doubt, but say there's still value in social priming's

ILLUSTRATION BY KAROL BANACH

central idea. It is worth studying whether it's possible to affect people's behaviour using subtle, low-cost interventions – as long as the more-outlandish and unsupported claims can be weeded out, says Esther Papies, a psychologist at the University of Glasgow, UK.

Equipped with more-rigorous statistical methods, researchers are finding that social-priming effects do exist, but seem to vary between people and are smaller than first thought, Papies says. She and others think that social priming might survive as a set of more modest, yet more rigorous, findings. "I'm quite optimistic about the field," she says.

Rise and fall

The roots of the priming phenomenon go back to the 1970s, when psychologists showed that people get faster at recognizing and processing words if they are primed by related ones. For instance, after seeing the word 'doctor', they recognized 'nurse' faster than they did unrelated words. This 'semantic' priming is now well established.

But in the 1980s and 1990s, researchers argued that priming could affect attitudes and behaviours. Priming individuals with words related to 'hostility' made them more likely to judge the actions of a character in a story as hostile, a 1979 study found. And in 1996, John Bargh, a psychologist at New York University in New York City found that people primed with words conventionally related to age in the United States – 'bingo', 'wrinkle', 'Florida' – walked more slowly than the control group as they left the lab, as if they were older⁵.

Dozens more studies followed, finding that priming could affect how people performed at general-knowledge quizzes, how generous they were or how hard they worked at

tasks. These behavioural examples became known as social priming, although the term is disputed because there is nothing obviously social about many of them. Others prefer 'behavioural priming' or 'automatic behaviour priming'.

In his 2011 best-seller *Thinking, Fast and Slow*, Nobel-prizewinning psychologist Daniel Kahneman mentioned several of the best-known priming studies. "Disbelief is not an option," he wrote of them. "The results are not made up, nor are they statistical flukes. You have no choice but to accept that the major conclusions of these studies are true."

But concerns were starting to surface. That same year, Daryl Bem, a social psychologist at Cornell University in Ithaca, New York, published a study suggesting that students could predict the future⁶. Bem's analysis relied on statistical techniques that psychologists regularly used. "I remember reading it and thinking 'Fuck. If we can do this, we have a problem,'" says Hans IJzerman, a social psychologist at the University of Grenoble Alps in Grenoble, France.

Also that year, three other researchers published a deliberately absurd finding: that those who listened to the Beatles song 'When I'm Sixty-Four' literally became younger than a control group that listened to a different song. They achieved this result by analysing their data in many different ways, getting a statistically significant result in one of them by simple fluke, and then not reporting the other attempts⁷. Such practices, they said, were common in psychology and allowed researchers to find whatever they wanted, given some noisy data and small sample sizes.

The papers had an explosive impact. Replication efforts that cast doubt on key findings

started to appear, including a 2012 report that repeated Bargh's ageing study and found no effect of priming unless the people observing the experiment were told what to expect⁸. It did not help that this all took place as it was discovered that a leading social psychologist in the Netherlands, Diederik Stapel, had been faking data for years.

In 2012, Kahneman wrote an open letter to Bargh and other "students of social priming", warning that "a train wreck" was approaching. Despite being a "general believer" in the research, Kahneman worried that fraud such as Stapel's, replication failures and a tendency for negative results not to get published had created "a storm of doubt" (see go.nature.com/2jcoetz).

Seven years later, the storm has uprooted many of social priming's flagship findings. Eric-Jan Wagenmakers, a psychologist at the University of Amsterdam, says that when he read the relevant part of Kahneman's book, "I was like, 'not one of these studies will replicate.' And so far, nothing has."

Psychologist Eugene Caruso reported in 2013 that reminding people of the concept of money made them more likely to endorse free-market capitalism⁹. Now at the University of California, Los Angeles, Caruso says that having tried bigger and more systematic tests of the effects, "there does not seem to be robust support for them". Ap Dijksterhuis, a researcher at Radboud University in Nijmegen, the Netherlands, says that his paper³ suggesting that students primed with the word 'professor' do better at quizzes "did not pass the test of time".

Kahneman told *Nature*: "I am not up to date on the most recent developments, so should not comment."

Researchers had been whispering about not being able to repeat big findings years before the priming bubble began to burst, says Nosek. Afterwards, in lessons shared with science's wider replication crisis, it became clear that many of the problematic findings were probably statistical noise – fluke results garnered from studies on too-small groups of people – rather than the result of fraud. It seems that many researchers were not alert to how easy it is to find significant-looking but spurious results in noisy data. This is especially so if researchers 'HARK' (Hypothesize After Results are Known) – that is, change their hypotheses after looking at their data. The fact that journals tend not to publish null results didn't help, because it meant the only findings that got through were the surprising ones.

There is also evidence that subconscious experimenter effects have been a problem, Papies says: one study found that when experimenters were aware of the priming effect they were looking for, they were much more likely to find it, suggesting that, subconsciously, they would affect the results in some way¹⁰.

WANING EFFECT

A meta-analysis of 246 experiments that exposed people to money-related stimuli found that early studies reported larger priming effects on behaviour, emotions and attitudes than did later ones. It also revealed larger effects in published work than in unpublished experiments provided by authors of the original studies.



Since then, there have been widespread moves throughout psychology to improve research methods. These include pre-registering study methods before looking at data, which prevents HARKing, and working with larger groups of volunteers. Nosek, for instance, has led the Many Labs project, in which undergraduates at dozens of labs try to replicate the same psychology studies, giving sample sizes of thousands. On average, about half of the papers that Many Labs looks at can be replicated successfully. Other collaborative efforts include the Psychological Science Accelerator, a network of labs that work together to replicate influential studies.

The new social priming

Today, much of the work being done in social priming involves replications of earlier work, or meta-analyses of multiple papers to try to tease out what still holds true. A meta-analysis of hundreds of studies on many kinds of money priming, reported this April¹¹, found little evidence for the large effects the early studies claimed. It also found larger effects in published studies than in unpublished experiments that had been shared with the authors of the meta-analysis (see ‘Waning effect’).

Original work hasn’t dried up entirely, says Papies, although the focus is changing. Much of the high-profile social-priming work of the past was designed to find huge, universal effects, she says. Instead, her group’s studies focus on finding smaller effects in the subset of people who already care about the thing being primed. She has found that people who want to become thinner are more likely to make healthy food choices if they are primed, say, with words on a menu such as ‘diet’, ‘thin’ and ‘trim figure’. But it works only in people for whom a healthy diet is a central goal; it doesn’t make everyone avoid fattening foods¹².

This matches the findings of a meta-analysis from 2015, led by psychologist Dolores Albarracín at the University of Illinois at Urbana-Champaign. It looked at 352 priming studies that involved presenting words to people, and it found evidence of real, if small, effects when the prime was related to a goal that the participants cared about¹³. That analysis, however, deliberately looked only at experiments in which the priming words were directly related to the claimed effect, such as rudeness-related words leading to ruder behaviour or attitudes. It avoided looking at studies with primes that had what it termed ‘metaphorical’ meaning – including the ageing-related words that Bargh said led to slower walking, or the money-related priming work.

Research into priming has declined, however, and what is considered priming is not always the same as the startling claims of the 1990s and 2000s. “There’s a lot less than there was five or ten years ago,” says Antonia Hamilton, a neuroscientist at University

College London, who still works on priming. Partly, she says, that’s because of the replication problems: “We do less since it all blew up. It’s harder to make people believe it and there are other topics that are easier to study.” It might also be simply that the topic has become less fashionable, she says.

Hamilton’s own work involves, among other things, putting people in functional magnetic resonance imaging (fMRI) scanners to see how priming affects brain activity. In one 2015 study, she used a scrambled-sentence task to prime ‘prosocial’ ideas (such as ‘helping’) and ‘antisocial’ ones (such as ‘annoying’), seeing whether it made participants quicker to mimic other people’s actions, and whether there were detectable differences in brain scans¹⁴.

Using fMRI is only practical with small numbers of volunteers, so she looks at how the same people respond when they have been primed and when they haven’t: a ‘within-subjects’ design, in contrast to the ‘between-subjects’ design of priming studies

“I still have no doubts whatsoever that in real life, behaviour priming works.”

that use a control group. The design means that researchers don’t have to worry about pre-existing differences between groups, Hamilton says. Her research has found priming effects: people primed with prosocial concepts behave in more prosocial ways, and fMRI scans did show differences in activity in brain areas such as the medial prefrontal cortex, which is involved in regulating social behaviours. But, she says, the effects are more modest than those the classic priming studies found.

Some researchers say that however efforts to test older results pan out, the concept of social or behavioural priming still has merit. “I still have no doubts whatsoever that in real life, behaviour priming works, despite the fact that in the old days, we didn’t study it properly relative to current standards,” says Dijksterhuis.

Bargh says that despite many researchers now discounting them, important early advances do exist – such as his own 2008 study, which reported that holding warm coffee made people behave more warmly towards others¹⁵. Direct replications have failed to support the result¹⁶, but Bargh says that a link between physical warmth and social warmth has been demonstrated in other work, including neuroimaging studies. “People say we should just throw out all the work before 2010, the work of people my age and older,” says Bargh, “and I don’t see how that’s justified.” He and Norbert Schwarz, a psychologist at the University of Southern California in Los Angeles, say that there have been replications of their earlier social-priming results – although critics counter that these were not

direct replications but ‘conceptual’ ones, in which researchers test a concept using related experimental set-ups.

Bargh says that results of social priming are still widely believed and used by non-academics, such as political campaigners and business marketers, even when they are sceptical. Gary Latham, for instance, an organizational psychologist at the University of Toronto in Canada, says: “I strongly disliked Bargh’s findings and wanted to show it doesn’t work.” Despite this, he says, he has for ten years consistently found that priming phone marketers with words related to ideas of success and winning increases the amount of money they make¹⁷. But Leif Nelson, a psychologist at the University of California, Berkeley, emphasizes that whether or not social-priming ideas are subsequently confirmed, the classic studies in the field were not statistically powerful enough to detect the things they claimed to find.

Bargh sees positives and negatives in how psychology research has changed. “If pre-registration stops people from HARKing, then I guess it’s good,” he says, “but it always struck me as an insult. ‘We don’t trust you to be honest’; it feels like we’re being treated like criminals, wearing ankle bracelets.”

Others disagree. The move towards open, reproducible science, according to most psychologists, has been a huge success. Social priming as a field might survive, but if it does not, then at least its high-profile problems have been crucial in forcing psychology to clean up its act. “I have to say I am pleasantly surprised by how far the field has come in eight years,” says Wagenmakers. “It’s been a complete change in how people do things and interpret things.”

Tom Chivers is a science journalist based in London.

1. Rohrer, D., Pashler, H. & Harris, C. R. *Basic Appl. Social Psychol.* **41**, 263–271 (2019).
2. Vohs, K. H., Meade, N. L. & Goode, M. R. J. *Science* **314**, 1154–1156 (2006).
3. Dijksterhuis, A. & van Knippenberg, A. J. *Pers. Soc. Psychol.* **74**, 865–877 (1998).
4. Lee, S. W. S. & Schwarz, N. J. *Pers. Soc. Psychol.* **103**, 737–749 (2012).
5. Bargh, J. A., Chen, M. & Burrows, L. J. *Pers. Soc. Psychol.* **71**, 230–244 (1996).
6. Bem, D. J. J. *Pers. Soc. Psychol.* **100**, 407–425 (2011).
7. Simmons, J. P., Nelson, L. D. & Simonsohn, U. *Psych. Sci.* **22**, 1359–1366 (2011).
8. Doyen, S., Klein, O., Pichon, C.-L. & Cleeremans, A. *PLoS ONE* **7**, e29081 (2012).
9. Caruso, E. M., Vohs, K. D., Baxter, B. & Waytz, A. J. *Exp. Psychol. Gen.* **142**, 301–306 (2013).
10. Gilder, T. S. E. & Heerey, E. A. *Psychol. Sci.* **29**, 403–417 (2018).
11. Lodder, P., Ong, H. H., Grasman, R. P. P. & Wicherts, J. M. J. *Exp. Psychol. Gen.* **148**, 688–712 (2019).
12. Papies, E. K. *Health Psychol. Rev.* **10**, 408–424 (2016).
13. Weingarten, E. et al. *Psychol. Bull.* **142**, 472–497 (2015).
14. Wang, Y. & Hamilton, A. *Soc. Cogn. Affect. Neurosci.* **10**, 486–493 (2015).
15. Williams, L. E. & Bargh, J. A. *Science* **322**, 606–607 (2008).
16. Chabris, C. F., Heck, P. R., Mandart, J., Benjamin, D. J. & Simons, D. J. *Soc. Psychol.* **50**, 127–132 (2019).
17. Stajkovic, A. D., Latham, G. P., Sergeant, K. & Peterson, S. J. *J. Bus. Psychol.* **34**, 791–802 (2018).

Books & arts



Pepper the robot greets visitors to Berlin's Futurium.

Lost in the house of tomorrow

What kind of future does Berlin's newest museum really show us? **By Stephen Cave**

Thirty years ago, the future became passé. When the Berlin Wall fell in late 1989 and the communist regimes that hid behind it collapsed, political scientist Francis Fukuyama called the event “the end of history”. But he also cast it as the finale of the future: the end of imagining how things might be different. The utopian visions driving both communism and fascism had been discredited and defeated. They were to be replaced by an eternal ‘now’ that, in Fukuyama’s words, saw “Western liberal democracy as the final form of human government”.

In the intervening decades, past and future have been reasserting themselves. Throwbacks such as right-wing populism are challenging liberal complacency, and environmental apocalypse is emerging as the face of the future. But most people in high-income countries would rather avert their eyes to focus on a still-comfortable consumerist present. At such a time, imagining a positive future is a challenge, but a new initiative in Berlin wants to try.

The Futurium, billed as a “house of futures”, is a museum, exhibition space and site for education and conferences. The €58-million (US\$64-million) building, designed by Berlin architects Richter Musikowski, shimmers in asymmetrical glass and metal, like a 1960s vision of the space age squashed between office blocks on the banks of the River Spree. It opened in September in the presence of Anja Karliczek, head of Germany’s Federal Ministry for Education and Research, the museum’s largest funder. Other backers include leading scientific institutes such as the Fraunhofer Society, Helmholtz Association and Max Planck Society.

How do these prestigious bodies invite us to imagine the future? Oddly timidly, in my view – with a focus on new technologies and modest hints for living sustainably. The opening exhibition reveals deeper issues with the thinking behind the space.

The Futurium’s three floors are divided into a basement ‘lab’ for workshops, a ground-floor ‘forum’ for larger events and an exhibition space above. The last is divided into three ‘thinking spaces’: Human, Nature and Technology, each with its own exhibition – respectively, Common Cause, Rethinking Nature and Towards New Horizons. A humanoid robot (‘Pepper’, made by SoftBank Robotics in Paris) acts as greeter, dispensing wristbands that contain implanted radio-frequency identification chips. At various points in the show, visitors can vote for a possible future by pressing

DAVID VON BECKER



Philip Beesley's cell-simulating 2018 sculpture *Noosphere* on display at the Futurium.

their wristband against a scanner. For example, they can choose whether they would prefer to live on after death as a digital avatar, pass their data to their loved ones or have all electronic traces of themselves deleted.

Past Pepper, the layout is baffling, with little indication of what is where, or why. When I stumbled across the Technology thinking space, however, I found the exhibits informative, with concepts such as big data and artificial intelligence explained well. Visitors of all ages are kept engaged by various gadgets, such as tablet computers that provide an augmented view of some displays.

Connected realms

The Human space invites us to consider topics such as the networked society and future cities. A central focus is overconsumption, with a rather messy series of exhibits prompting visitors to consider how much stuff they really need, and promoting the benefits of decluttering. In the Nature area, themes include urban farming and renewable energy. An exhibit on sustainable biomaterials presents some provocative alternatives to plastics — shoes made from algae, and fungus-based furniture. The curators intend

to update the whole exhibition gradually, and to introduce major new themes annually from 2021. These individual elements are stimulating.

But I found the conceptual basis for the Futurium's structure — the division into human, technology and nature — problematic. An installation at the start of the exhibition displays the three concepts as linked but

“To portray technology as unconnected to humans fuels technological determinism.”

distinct, with ‘technology’ far removed from ‘people’, and ‘natural’ nodes such as ecosystems or even the laws of nature on a par with art and faith. These separations are deeply misleading. To portray technology as unconnected to humans fuels technological determinism; to portray it as separate from nature is to perpetuate the Western myth that technology exists to dominate an unruly Earth.

Prising the human away from nature is perhaps the most egregious of all; it implies

that we are self-contained, rather than dependent on a biosphere increasingly under threat. Other framings, such as humanity as a subset of nature, and technology as a subset of the human, might have been better suited to the messages about sustainability that the museum ultimately wants to promote.

Overall, the Futurium succeeds best as a showcase for the shiniest aspects of the present. In this way, it resembles other tech-engagement centres, such as Science Gallery Dublin and its six sister venues around the world, or Tokyo's National Museum of Emerging Science and Innovation. But it claims to be something more: a place for co-imagining alternative futures. To succeed, it will need to be bolder. Even though the Berlin landscape is dotted with monuments to failed ideologies, such as the Stasi Museum, history did not end when the wall fell. To imagine new futures, this museum must free itself from the conceptual frameworks of the past.

Stephen Cave is executive director of the Leverhulme Centre for the Future of Intelligence at the University of Cambridge, UK.
e-mail: sjc53@cam.ac.uk

CRISPR: the movie

A documentary showcases biology's hottest tool – up to the point when things went awry. **By Amy Maxmen**

At the start of *Human Nature*, a documentary about the gene-editing tool CRISPR, we meet a young man with sickle-cell anaemia. David Sanchez is wise beyond his years, driving home the injustice of his gruelling blood infusions and shortened lifespan. Researchers are testing a therapy for his condition in clinical trials using CRISPR.

This is a film probing the unknown future of a technology that, in the past decade, has skyrocketed from obscurity to become the subject of a Netflix series called *Unnatural Selection*, which debuted on 18 October (the trailer promises provocation by leading with biohackers injecting the editing tool). *Human Nature* does not take a shock-horror approach. This is the film that scientists would probably prefer the public to see.

The project began with a meeting between the Wonder Collaborative, a scientific-documentary organization based in San Francisco, California, and CRISPR co-discoverer Jennifer Doudna, a biochemist at the University of California, Berkeley, and her colleagues. The scientists guided the film-makers, led by a team of co-producers (including former cell biologist Sarah Goodwin and journalists Dan Rather and Elliot Kirschner) and director Adam Bolt, on the scientific and ethical issues. And the film-makers read up on the technology themselves. When the film was nearly complete, they sought feedback from members of the US National Academies of Sciences, Engineering, and Medicine.

As a result, the film is high on the thrill and potential of discovery, and every scientist and bioethicist featured is passionate and thoughtful. Notably, the documentary sidesteps academic politics, such as the ongoing patent battle over who will reap CRISPR's financial rewards. And it lacks input from scientists outside the United States and western Europe – such as from China, where the first human embryos have been edited – or from policy-makers with the power to restrict what research is permissible. However, it does feature a handful of people who have genetic disorders, as well as parents of children with these maladies.

Straight to the point

Human Nature follows a straight path, beginning with the seeds of genetic engineering in the 1960s and ending in the ethical quagmires of making changes to human DNA that

Human Nature

Director: Adam Bolt (2019)
<https://wondercollaborative.org/>

future generations could inherit. The film will be educational for people who haven't heard of CRISPR before; little will be new for those who have. Having covered the story from the beginning, I enjoyed watching the who's who of CRISPR's early years speak for themselves.

And the technology itself is covered beautifully. The camera swoops over crystalline salt pools in Spain, where Francisco Mojica, a microbiologist at the University of Alicante, ponders the repeated genetic letters in bacterial genomes – and the 'spacers' of unidentified DNA sequences in between. In 2005, he revealed that these enigmatic spacers mimic genes from viruses that once infected the microbes, and that these sequences form a kind of 'memory' that allows bacteria to recognize and attack the invaders in future.

Animations of double helices clarify what happens; they reappear when Doudna's team adapts CRISPR as a laboratory workhorse. Even this scene has a pulse, owing to the ever-enthusiastic Fyodor Urnov, science director at the Innovative Genomics Institute, Berkeley's centre focused on CRISPR. Together with the enzyme Cas9, CRISPR spots and snips genes that it's been programmed to find.

Feng Zhang at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, appears briefly. But scant attention is devoted to his groundbreaking studies from 2013, which

enabled CRISPR to edit animal DNA (this is the work at the centre of the Broad's contested patents). And although a montage of headlines shows a subsequent flurry of experiments in which scientists edit various organisms, little is said about how these studies have pushed forward fields ranging from evolutionary biology to agriculture. Nor do we learn about scientists' continuing struggle to use CRISPR to insert genes in many organisms other than mice and flies. Their attempts often fail, or result in side effects or death.

The film-makers probably feared that audiences would grow weary of details. But I was hungry for an update on what scientists are actually doing right now. That's especially true when it comes to editing people. But rather than dwell on a couple of dozen CRISPR-based therapies in the early stages of testing, the film bounds into the possibilities of engineering humans using CRISPR with an ominous clip of Russian President Vladimir Putin speaking in 2017. Soldiers, he says, could be endowed with the ability to fight without pain. And Urnov explains how: CRISPR could be used to delete a gene that transmits pain signals to the brain. He says he's sure that this will become an analgesic treatment offered to cancer patients in pain.

Ethical dilemma

The rest of the film centres on the ethics of human editing, as if everything were possible. George Daley, a stem-cell biologist at Harvard Medical School in Boston, Massachusetts, draws a line between editing embryos and editing adults; in the former case, children would be born with the alteration in all their cells, and changes would be passed down through generations.

All the scientists (the protagonists of this film) are conscientious. Doudna, for example, says that Adolf Hitler showed up in one of the nightmares she had that were triggered by her anxiety over the potential misuse of the tool. In 2015, Urnov and his colleagues argued that



David Sanchez has sickle-cell anaemia, which could one day be addressed using gene editing.

WONDER COLLABORATIVE

editing embryos, sperm and eggs should be banned for now (see E. Lanphier *et al. Nature* **519**, 410–411; 2015). And earlier this year, Zhang and others recommended that scientists come up with a framework that governments could use to evaluate research proposals as the science of gene editing progresses (see E. Lander *et al. Nature* **567**, 165–168; 2019).

Still, other scientists can't contain their excitement. Stephen Hsu, a co-founder of Genomic Prediction, a genetic-testing company for *in vitro* fertilization in North Brunswick, New Jersey, suggests that, eventually, editing could advance humanity by making people healthier, longer-lived or more intelligent. The documentary pushes him on this point, and flashes a Nazi propaganda clip. Hsu counters that his vision is different from eugenics because the choice to edit is made by parents.

Troubling development

Alta Charo, a bioethicist at the University of Wisconsin–Madison, also dismisses certain fears, pointing out, for example, that characteristics such as intelligence are controlled by multiple genes and by the environment. But she concedes that there is a risk to editing, and therefore it shouldn't be used frivolously. Just 30 out of 195 countries have banned the editing of human embryos, sperm and eggs in the clinic with CRISPR, and the rules might not govern pure research.

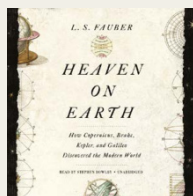
Human Nature traces CRISPR up to a pivotal moment. The film was nearly finished when the news was reported last November that twin girls had been born after Chinese biophysicist He Jiankui had used CRISPR to edit their embryos. So the film-makers just spliced this development in as an afterthought. After the heroes of the film had spent so much time expounding on the need to prevent this outcome, its sudden fruition is troubling.

Asked why the film-makers didn't revise the documentary to focus on the case, Kirschner says that they decided there was value in what they had: a film on CRISPR's origins. Plus, we truly don't know what will happen next. Kirschner writes: "It is impossible to tell whether it will ultimately be seen as an inflection point or an aberration."

The Wonder Collaborative had considered creating just a brief CRISPR explainer. I'm glad the group opted for a full-length feature: it gives them time to strike a nerve. For me, this happened in the scenes with Sanchez. At the end of the film, after so many researchers have gushed about the power of CRISPR to cure disease, the interviewer asks Sanchez if he wished his parents had used the tool to prevent his being born with a deadly condition. Sanchez pauses, and says no: "I don't think I'd be me."

Amy Maxmen writes for *Nature* from Oakland, California.

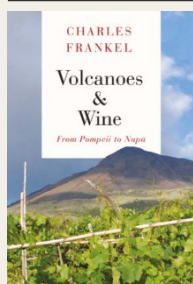
Books in brief



Heaven on Earth

J. S. Fauber *Pegasus* (2019)

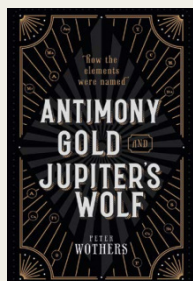
Four towering sixteenth-century scientists — Nicolaus Copernicus, Tycho Brahe, Johannes Kepler and Galileo Galilei — discovered heliocentrism at a time of sociopolitical tumult. As J. S. Fauber drives home in this dynamic science history, their intermeshed stories form a mighty "intergenerational epic" sweeping in the likes of Brahe's sister Sophie and Galileo's daughter Virginia. A wonderfully wrought explication of how a powerful thesis began its journey to becoming unavoidable fact, and seeded modernity in the process.



Volcanoes and Wine

Charles Frankel *Chicago University Press* (2019)

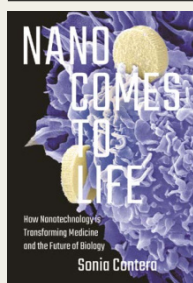
For this intriguing exploration of volcanism and viticulture, Charles Frankel scoured geologically active regions to trace how soils and landforms shape local wines. He begins with the 1620 BC eruption on Santorini that left the Greek island little more than a caldera, yet created ideal conditions for growing Assyrtiko grapes, used in unctuous Vinsanto. No less gripping are Frankel's descriptions of the deep-time lava flows and flooding that formed Oregon's Pinot-growing Willamette Valley. A gem for geologists and wine buffs alike.



Antimony, Gold and Jupiter's Wolf

Peter Wothers *Oxford University Press* (2019)

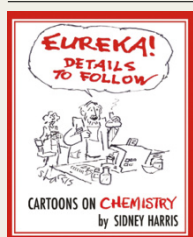
Hydrogen, caesium, silver: how were elements named? In this stimulating chemical chronicle, Peter Wothers unravels tangled etymologies. Eighteenth-century chemist Antoine Lavoisier, for instance, named oxygen to signify 'acid-former', only to have the word construed as "the son of a vinegar merchant". W, the symbol for tungsten, is a nod to its traditional moniker wolfram (derived from 'wolf-foam'). From copper to californium, we discover how the sober face of the periodic table hides dramatic backstories.



Nano Comes to Life

Sonia Contera *Princeton University Press* (2019)

Nanotechnology researcher Sonia Contera's succinct study surveys the progress of nano-tools in biological and medical research. As she relates, there is much in development: DNA technology aimed at crafting nanoscale machines to target specific cancer cells; nano-antibiotics for fighting infection; and nano-approaches to tissue engineering. Contera frames this near-future transmaterial science, with its focus on human well-being, as an effort allied to social justice even as it probes existential questions of what it means to be human.



Eureka! Details to Follow

Sidney Harris *Science Cartoons Plus* (2019)

After a year of bouts at the bench (and blasts of bad news), you may need relief. Science cartoonist Sidney Harris — whose work has graced the pages of many journals, including *Nature*, offers an ace antidote in this irreverent look at chemistry. Here is Lewis Carroll's Alice, thwarted by a looking-glass made of Lexan polycarbonate resin; two chemists absorbing the news that "you can't both be the 'father' of ammonium pentoxide phosphate"; the Institute for Advanced Hindsight; and oodles more. **Barbara Kiser**

Margarita Salas

(1938–2019)

Biochemist whose discoveries led to faster, more-accurate DNA testing.

Margarita Salas discovered a new mechanism for the replication of DNA. The enzyme she isolated as the key to it has transformed the process of amplifying DNA from very small samples, and is now widely used in forensics, studies of ancient DNA and oncology, as well as in basic research. Her invention, the most profitable patent ever filed by the Spanish National Research Council (CSIC), was recognized earlier this year by a lifetime achievement award from the European Patent Office.

Salas, who died on 7 November, aged 80, was central in bringing modern molecular biology to Spain. Coming of age during the regime of General Francisco Franco, who denied women an equal role in the workplace, she blazed a path for female scientists in her country. She was one of the first winners of the L'Oréal–UNESCO awards for women in science, and in 2008, King Juan Carlos I awarded her the hereditary title of marquise, in recognition of her impact on Spanish society.

Salas was born in Canero, a small village in the Asturias region in the north of Spain, in 1938, in the middle of the Spanish Civil War. Her mother was a schoolteacher; her father, a physician, influenced her interest in science. In 1955, she went to Madrid University to study chemistry. In 1958, on a visit home, she had an inspirational meeting with a distant relative – the biochemist Severo Ochoa, who was awarded a Nobel prize the following year. Fascinated by the emerging discoveries in biochemistry, Salas did a PhD on yeast metabolism in the laboratory of enzymologist Alberto Sols in Madrid. There she met and married the biochemist Eladio Viñuela.

In 1964, Salas and her husband were accepted as postdoctoral fellows in Ochoa's lab at New York University. The lab was deciphering the mechanisms that transfer genetic information from DNA to protein – replication, transcription and translation. Salas contributed two key pieces of information. First, she found that these mechanisms read the linear DNA message in only one direction. Second, she helped to show that the triplet of RNA nucleotides UAA represents a stop codon, where protein synthesis ends.

Returning to Spain in 1967, Salas and Viñuela set up a lab at the CSIC's Centre for Biological Research in Madrid. They carried out a systematic genetic analysis of a bacterial



virus, the phage $\phi 29$. Salas discovered that the linear, double-stranded DNA of the virus had a specific protein attached at one end of each strand. She found that this protein acts as a starting point when the viral genome replicates. The mechanism also operates in some mammalian viruses, including those that infect humans, such as adenovirus and hepatitis B virus.

“For generations of Spanish scientists, she was a guiding light.”

In 1977, Salas made her last academic move – to the recently founded Severo Ochoa Centre for Molecular Biology (CBMSO). The centre brought together most of the leading research groups in molecular biology in Spain in a joint institution of the CSIC and the Autonomous University of Madrid. Here, with her colleague Luis Blanco, Salas isolated the DNA polymerase enzyme from $\phi 29$ that went on to have such a profound impact on DNA analysis.

Existing methods for DNA amplification, using the polymerase chain reaction (PCR), were slow and prone to errors. Using $\phi 29$ DNA polymerase proved to be much more efficient: the technique was called multiple displacement amplification. It was highly accurate,

able to generate very large DNA fragments, and it worked at a constant temperature in the lab, avoiding the need for shifts in temperature as in PCR. Requiring only very small samples to generate many copies of whole genomes, it proved ideal for forensic analysis, for the identification of mutations in tumours and for genetic analysis of traces of DNA from ancient bones.

Salas and Blanco patented the enzyme, and its commercial licensing has provided an important source of income for the CSIC. Devoted to basic research, Salas was also delighted that her work had led to this important application. She was director of the CBMSO from 1992 to 1994, and continued to work enthusiastically at experiments in her lab until a few weeks before her death.

Salas was a courageous role model. She was the first woman to chair a big biomedical research centre, to be president of the Spanish Biochemistry and Molecular Biology Society, and the first to be elected to several academies.

After receiving awards or when celebrating an anniversary, Margarita, who was a shy person with an austere lifestyle, would invite her many current and former students to dinner at a smart restaurant. Here, she would talk informally, about topics unrelated to science, from Bach's music to modern Spanish literature, revealing her wide interest in culture.

Margarita conveyed her passion for scientific research and the thrill of discovery to her students, fostering their motivation, creativity, rigour and perseverance. She created a true school, teaching molecular biologists from different places and of different origins how to conduct – and take joy in – research. Many of her trainees are now research leaders.

For generations of Spanish scientists, she was a guiding light. We will miss her, her mentorship and her leadership.

Jesús Avila, Federico Mayor Jr and Lourdes Ruiz-Desviat are at the Severo Ochoa Centre for Molecular Biology, Madrid, Spain. **J.A.** was Margarita Salas' first PhD student when she returned to Spain, and is currently the director of CIBERNED, Spain. **F.M.Jr** was a former director of the Severo Ochoa Centre for Molecular Biology and **L.R.-D.** is the current director.
e-mails: jesus.avila@csic.es; fmayor@cbm.csic.es; lruiz@cbm.csic.es

FRANCIS TSANG/COVER/GETTY

Comment



ILLUSTRATION BY DAVID PARKINS

Predatory journals: no definition, no defence

Agnes Grudniewicz, David Moher, Kelly D. Cobey and 32 co-authors

Leading scholars and publishers from ten countries have agreed a definition of predatory publishing that can protect scholarship. It took 12 hours of discussion, 18 questions and 3 rounds to reach.

When 'Jane' turned to alternative medicine, she had already exhausted radiotherapy, chemotherapy and other standard treatments for breast cancer. Her alternative-medicine practitioner shared an article about a therapy involving vitamin infusions. To her and her practitioner, it seemed to be authentic grounds for hope. But when Jane showed the article to her son-in-law (one of the authors of this Comment), he realized it came from a predatory journal – meaning its

promise was doubtful and its validity unlikely to have been vetted.

Predatory journals are a global threat. They accept articles for publication – along with authors' fees – without performing promised quality checks for issues such as plagiarism or ethical approval. Naïve readers are not the only victims. Many researchers have been duped into submitting to predatory journals, in which their work can be overlooked. One study that focused on 46,000 researchers based in Italy found that about 5% of them published in such outlets¹. A separate analysis suggests predatory publishers collect millions of dollars in publication fees that are ultimately paid out by funders such as the US National Institutes of Health (NIH)².

One barrier to combating predatory publishing is, in our view, the lack of an agreed definition. By analogy, consider the historical criteria for deciding whether an abnormal bulge in the aorta, the largest artery in the body, could be deemed an aneurysm – a dangerous

condition. One accepted definition was based on population norms, another on the size of the bulge relative to the aorta and a third on an absolute measure of aorta width. Prevalence varied fourfold depending on the definition used. This complicated efforts to assess risk and interventions, and created uncertainty about who should be offered a high-risk operation³.

Everyone agrees that predatory publishers sow confusion, promote shoddy scholarship and waste resources. What is needed is consensus on a definition of predatory journals. This would provide a reference point for research into their prevalence and influence, and would help in crafting coherent interventions.

To hammer out such a consensus and to map solutions, we and others met in Ottawa, Canada, over two days in April this year. The 43 participants hailed from 10 countries and represented publishing societies, research funders, researchers, policymakers, academic institutions, libraries and patient partners (that is, patients and caregivers who proactively engage in research). Our focus was the biomedical sciences, but our recommendations should apply broadly.

Here we put forward our definition. We describe what it took to achieve consensus and how we'll move forward.

The definition

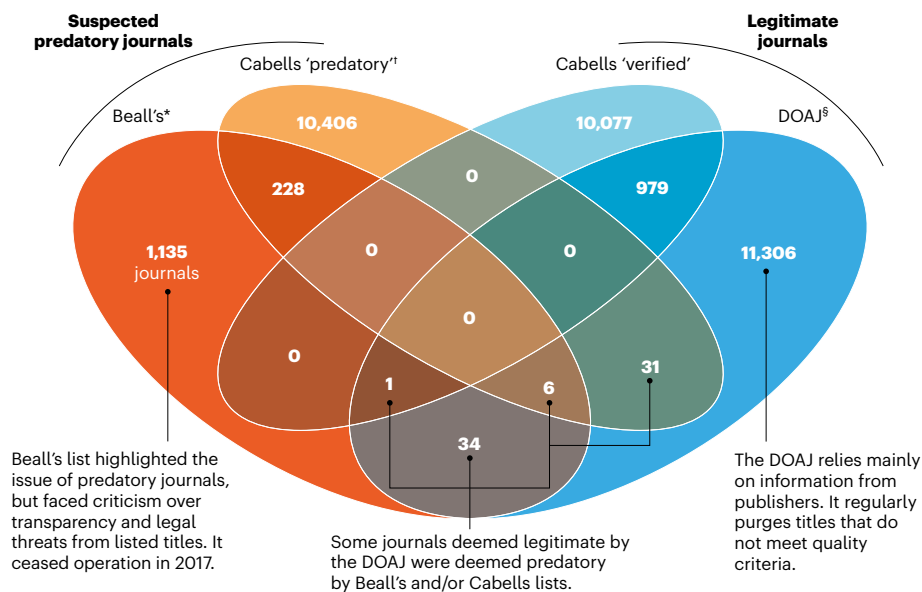
The consensus definition reached was: "Predatory journals and publishers are entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices."

Road to consensus

Since the term 'predatory publishers' was coined in 2010, hundreds of scholarly articles, including 38 research papers, have been written warning about them. Scientific societies and publishers (including Springer Nature) have helped to establish the 'Think. Check. Submit.' campaign to guide authors. But it is not enough. More than 90 checklists exist to help identify predatory journals using characteristics such as sloppy presentation or titles that include words such as 'international'. This is an overwhelming number for authors. Only three of the lists were developed using research evidence⁴. Paywalled lists of quality journals and predatory journals show that there is an appetite for clear, authoritative guidance. But these lists are inconsistent and sometimes out of reach^{5,6} (see 'No list to rule them all').

NO LIST TO RULE THEM ALL

Assessments of which journals are likely to be predatory or legitimate do not tally, and titles can appear in both categories. There is no way to know which journals were considered for a list but left off, or which were not considered.



⁴Informally assessed by University of Colorado Denver librarian Jeffrey Beall in ~2008-17; ⁵Pay-to-access lists from Cabells, a scholarly analytics company; ⁶The Directory of Open Access Journals, a community-curated list requiring journal best practices such as peer review and statements on author fees and licensing.

A journal's membership of agencies such as COPE (the Committee on Publication Ethics), curated indexes such as Web of Science, or being listed in the Directory of Open Access Journals (DOAJ) is insufficient to guarantee quality. Predatory journals have found ways to penetrate these lists, and new journals have to publish for at least a year before they can apply for indexing.

A scoping review comparing publications about predatory journals found that their characterizations sometimes overlapped, sometimes did not and sometimes directly conflicted⁷. These inconsistencies suggest that crafting a practical definition would require building consensus across researchers, publishers, research institutions and the broader public.

Participants in our summit completed a three-round modified Delphi survey (a structured technique to elicit input, offer feedback and build consensus) that included 18 questions and 28 sub-questions. There were also 12 hours of discussion, followed by 2 further rounds of feedback and revision.

Crafting a consensus definition was hard. Even reaching agreement on the use of 'predatory' was a challenge. Part of the group wanted a term that acknowledges that some authors turn to these outlets fully aware of their low quality; these scholars willingly pay to publish in predatory journals to add a line to their

CVs. We discussed replacing the term entirely with language that recognizes nuances in publishers' quality and motivation. Alternatives considered included 'dark', 'deceptive', 'illegitimate' and 'acting in bad faith'. Ultimately, we concluded that the term 'predatory' has become recognized in the scholarly community. Implementation science suggests that introducing new nomenclature would take considerable resources, which we felt could be better put towards combating predatory publishing directly. So we recommend keeping the word 'predatory' while noting its limitations.

Details matter

Predatory journals are driven by self-interest, usually financial, at the expense of scholarship. They are characterized by the following:

False or misleading information. This applies to how the publisher presents itself. A predatory journal's website or e-mails often present contradictory statements, fake impact factors, incorrect addresses, misrepresentations of the editorial board, false claims of indexing or membership of associations and misleading claims about the rigour of peer review.

Deviation from best editorial and publication practices. Standards here have been set out in the joint statement on Principles of Transparency and Best Practice in Scholarly

Comment

Publishing (see go.nature.com/35mq7mj), issued by the DOAJ, the Open Access Scholarly Publishers Association, COPE and the World Association of Medical Editors. Examples of substandard practice include not having a retraction policy, requesting a transfer of copyright when publishing an open-access article and not specifying a Creative Commons licence in an open-access journal. These characteristics can be difficult to know before submitting, although such information is easily obtained from legitimate journals. An unprofessional-looking web page – with spelling or grammar mistakes or irrelevant text – should also raise red flags.

Warning signs should be assessed with care. For instance, journals are not eligible for listing on the DOAJ or joining COPE until after one year of operation. A well-meaning but poorly resourced journal might not be able to maintain a professional website. Also, some journals claim to follow best practice but do not. Summit participants agreed that the burden of proof rests on the journal.

Lack of transparency. There are two reasons we list this separately from deviation from best practice. First, transparency in operational procedures (such as how editorial decisions are made, fees applied and peer review organized) is presently somewhat aspirational in academic publishing and thus cannot be considered a current best practice. Second, the absence of transparency in predatory journals makes it important enough to highlight separately. Predatory publishers often fail to provide their contact information or details about article processing charges. Editors and members of their editorial boards are often unverifiable.

Aggressive, indiscriminate solicitation. Although legitimate journals might solicit submissions, predatory journals often use aggressive solicitation such as repeated e-mails. These might be excessively flattering in tone, or might mention researchers' past publications while noting that related submissions are urgently needed for a forthcoming issue. A clear warning sign is that the invitee's expertise is outside the journal's scope.

Criteria we left out. Some obvious candidates for this list – journal quality and intent to deceive – were deliberately left out. It can be tough to distinguish a predatory journal from a journal that is under-resourced. Both can be low quality, but the latter does not have an intention to deceive^{8,9} (see also go.nature.com/33gmjnt and go.nature.com/2afaka7). Furthermore, such intent is hard to assess and, if many of the characteristics described in the definition are met, identifying intent might not be necessary.

Most controversially, we omitted quality

of peer review, even though negligent peer review is often a prominent feature of predatory journals. We are not saying that peer review is unimportant, only that it is currently impossible to assess. Unfortunately, many legitimate journals fail to make their peer-review processes sufficiently transparent, for instance by sharing peer reviewers' comments and other data. At the moment, journal quality, adequacy of peer review and deceit are too subjective to include.

Next steps

Efforts to fight predatory publishing require collaboration and support. Organizations, researchers and governments have started to respond. To name just a few, in 2017, the NIH released a statement encouraging researchers it funds to publish in reputable journals. India's University Grants Commission has created a reference list of respectable journals and is currently working to revise academic publication incentives and develop a training course to reinforce the message. In November 2018, COPE held a forum on predatory publishing to examine problems and solutions.

So far, disparate attempts to address predatory publishing have been unable to control this ever-multiplying problem. The need will be greater as authors adjust to Plan S and

"It can be tough to distinguish a predatory journal from a journal that is under-resourced."

other similar mandates, which will require researchers to publish their work in open-access journals or platforms if they are funded by most European agencies, the World Health Organization, the Bill & Melinda Gates Foundation and others (see www.coalition-s.org).

Many might argue that, with predatory journals adapting so quickly, our group's efforts would have been better spent crafting interventions or promoting outreach. We believe that with this consensus definition, we are better prepared to track the problem over time, compare the results of studies on predatory journals and develop and evaluate intervention strategies such as educational campaigns and policy mandates. Over the coming months, we will solicit input and make the definition usable so that funders and academic institutions can ensure that researchers avoid submitting manuscripts to predatory journals or listing such publications on their CVs.

Our first step is to develop a portal that presents our definition and other educational resources in multiple languages – available at <https://osf.io/8xvpm> – and how to get involved. Next, we will establish an international observatory to compile data on the

problem, tracking numbers of publications in predatory journals by discipline and geography. We will work with funders, institutions, patients and other stakeholders to iteratively develop resources to assess journal quality. We are seeking funding to create and test a digital tool to achieve these goals.

Efforts to counter predatory publishing need to be constant and adaptable. The threat is unlikely to disappear as long as universities use how many publications a scholar has produced as a criterion for graduation or career advancement. The publish-or-perish culture, a lack of awareness of predatory publishing and difficulty in discerning legitimate from illegitimate publications fosters an environment for predatory publications to exist. Predatory journals are also quick to adapt to policies and measures designed to foil them. As scientific publishers experiment with new formats and business models online, it has become increasingly easy for fake publishers to masquerade as legitimate ones. We invite others to join us in our call to action.

The authors

Agnes Grudniewicz is an assistant professor at the Telfer School of Management, University of Ottawa, Canada. **David Moher** is a senior scientist and director and **Kelly D. Cobey** is an investigator at the Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada.

Gregory L. Bryson, Samantha Cukier, Kristiann Allen, Clare Arden, Lesley Balcom, Tiago Barros, Monica Berger, Jairo Buitrago, Lucia Cugusi, Michael R. Donaldson, Matthias Egger, Ian D. Graham, Matt Hodgkinson, Karim M. Khan, Mahlubi Mabizela, Andrea Manca, Katrin Milzow, Johann Mouton, Marvelous Muchenje, Tom Olijhoek, Alexander Ommaya, Bhusan Patwardhan, Deborah Poff, Laurie Proulx, Marc Rodger, Anna Severin, Michaela Strinzel, Mauro Sylos-Labini, Robyn Tamblyn, Marthie van Niekerk, Jelte M. Wicherts, Manoj M. Lal.

e-mails: agnes.grudniewicz@telfer.uottawa.ca; dmoher@ohri.ca; kcobey@ohri.ca

1. Bagues, M., Sylos-Labini, M. & Zinovyeva, N. *Res. Pol.* **48**, 462–477 (2019).
2. Moher, D. et al. *Nature* **549**, 23–25 (2017).
3. Moher, D., Cole, C. W. & Hill, G. B. *Eur. J. Vasc. Surg.* **6**, 647–650 (1992).
4. Cukier, S. et al. Preprint at medRxiv <https://doi.org/10.1101/19005728> (2019).
5. Strinzel, M., Severin, A., Milzow, K. & Egger, M. *mBio* **10**, e00411-19 (2019).
6. Teixeira da Silva, J. A. & Tsigaris, P. *J. Acad. Librariansh.* **44**, 781–792 (2018).
7. Cobey, K. D. et al. *F1000Res* **7**, 1001 (2018).
8. Berger, M. & Cirasella, J. *C&RL News* **76**, 132–135 (2015).
9. Eriksson, S. & Helgesson, G. *Learn. Publ.* **31**, 181–183 (2018).

A full list of author affiliations accompanies this Comment online (see go.nature.com/2dndsfj).

Correspondence

Groundwater: a call to action

As we embark on the United Nations 'decade of action' (see go.nature.com/2opvyi3), and as this week's UN COP25 Climate Change Conference concludes in Madrid, let's remember the crucial contribution of groundwater to climate resilience and sustainable development.

Besides sustaining drinking water and ecosystems worldwide, groundwater acts as a subsurface sponge for floods. It is a resource against drought and for natural climate solutions that sequester soil carbon. And it is crucial for sustainable development because it enables food security and lifts rural populations out of poverty.

However, these essential benefits are being undermined by the long-term depletion, contamination and salinization of groundwater (see, for example, I. E. M. de Graaf *et al.* *Nature* **574**, 90–94; 2019).

In our view, groundwater needs to be monitored and managed with greater rigour on regional and global scales so that it can be used more effectively to boost climate adaptation and sustainable development. As members of a global group of scientists and practitioners, we have issued a call to action to international and national governmental and non-governmental agencies, development organizations, corporations, decision makers and scientists, to ensure that groundwater benefits society now and into the future (see go.nature.com/37gnbtb).

Tom Gleeson* University of Victoria, British Columbia, Canada.

tgleeson@uvic.ca

*On behalf of 5 correspondents; see go.nature.com/2dfulej

Evaluating Italy's ranking boom

The president and vice-president of the Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) claim that Italy's rise in international research-impact rankings is a real effect (P. Miccoli and R. I. Rumiat *Nature* **574**, 486; 2019), and not (as we have argued) the result of Italian scholars citing one another's articles more heavily (see *Nature* <http://doi.org/dcgj>; 2019). We question their evidence for this claim.

First, they say that scientific productivity in Italy has risen in the past decade, possibly stimulated by the introduction of performance-related university funding. More articles are indeed being published, but the yearly growth rate of Italy's scientific production has in fact slowed down since the introduction of performance-related targets in 2012, according to ANVUR's own statistics (see go.nature.com/34ms9n; in Italian).

Second, they state that ANVUR recognizes the importance of correcting gaming behaviours, including self-citation. They point out that, in an evaluation of 2011–14 work, the agency established a criterion for 'downgrading' papers in which self-citation exceeded a given threshold. ANVUR's own reports, however, show that this downgrading was never applied (see go.nature.com/2jn2si; in Italian).

In our view, ANVUR's claim needs to be grounded more in fact and less in aspiration.

Alberto Baccini, Eugenio Petrovich University of Siena, Italy.

baccini@unisi.it

Giuseppe De Nicolao University of Pavia, Italy.

We'll take 'quantum advantage'

We take issue with the use of 'supremacy' when referring to quantum computers that can out-calculate even the fastest supercomputers (F. Arute *et al.* *Nature* **574**, 505–510; 2019). We consider it irresponsible to override the historical context of this descriptor, which risks sustaining divisions in race, gender and class. We call for the community to use 'quantum advantage' instead.

The community claims that quantum supremacy is a technical term with a specified meaning. However, any technical justification for this descriptor could get swamped as it enters the public arena after the intense media coverage of the past few months.

In our view, 'supremacy' has overtones of violence, neocolonialism and racism through its association with 'white supremacy'. Inherently violent language has crept into other branches of science as well – in human and robotic spaceflight, for example, terms such as 'conquest', 'colonization' and 'settlement' evoke the *terra nullius* arguments of settler colonialism and must be contextualized against ongoing issues of neocolonialism.

Instead, quantum computing should be an open arena and an inspiration for a new generation of scientists.

Carmen Palacios-Berraquero Nu Quantum, Cavendish Laboratory, Cambridge, UK.
carmen@nu-quantum.com

Leonie Mueck Riverlane, Cambridge, UK.

Divya M. Persaud University College London, UK.

Supported by 13 signatories; see go.nature.com/2yuLvs

Proven protection against air pollution

As researchers for the certifying body for personal protective equipment in the United States, we caution against misinterpretation of Wei Huang's and Lidia Morawska's contention that face masks could increase health risks from air pollution (*Nature* **574**, 29–30; 2019).

Although the authors attempt to distinguish between 'medical masks' and 'specialist respirators', a clearer definition of 'mask' would avoid confusion over the capabilities of different protective devices. As they point out, surgical masks are loose-fitting and ineffective against air pollution. However, respirators approved by the National Institute for Occupational Safety and Health (NIOSH) fit tightly to the face and filter at least 95% of airborne particles, including aerosolized nanoparticulates (E. Vo *et al.* *Ann. Occup. Hyg.* **59**, 1012–1021; 2015).

Even NIOSH-approved respirators that have not been personally fitted provide some protection in non-occupational settings (see go.nature.com/35ztfy). Outdoor workers in California, for example, wore such devices as safeguards against non-oily particulate hazards produced by this year's wildfires (go.nature.com/35jwdw).

An absence of evidence from clinical trials is no reason not to take precautionary measures.

Tyler D. Quinn, Lew Radonovich, Maryann D'Alessandro Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Pittsburgh, Pennsylvania, USA.
yhh7@cdc.gov

Disclaimer: These conclusions do not necessarily represent the official position of NIOSH.

News & views

Medical research

Ready-made cellular plugs heal skin wounds

Mark C. Coles & Christopher D. Buckley

The finding that a thin sheet of fibrous tissue under the skin contains a prefabricated, movable cellular sealant that can heal deep wounds might have implications for the treatment of scars and ulcers. **See p.287**

Skin consists of an outer epidermal layer (the epidermis) and an inner dermal layer (the dermis). If you pinch your skin, you can lift it because these two cellular layers move freely above a membranous sheet called the fascia, which contains cells and extracellular-matrix material. This gelatinous tissue creates a frictionless interface between the skin and the more rigid structures beneath it, such as muscle and bone. However, it now seems that the fascia has roles beyond providing a non-stick surface. On page 287, Correa-Gallegos *et al.*¹ report that the fascia contains a movable sealant that patches up deep injuries to enable rapid wound repair.

The scar tissue of a healing skin wound contains fibroblast cells, which make and modify extracellular-matrix proteins. These fibroblasts can be identified by their expression of a protein called Engrailed-1, and are termed Engrailed-positive fibroblasts (EPFs). The idea that the fascia might be a repository of cellular components involved in wound healing and scar formation came from a previous study², which reported that EPFs reside not only in the skin, as expected, but also in the fascia.

To investigate wound healing in mice, Correa-Gallegos and colleagues grafted fascia that contained cells engineered to express green fluorescent protein onto skin cells expressing red fluorescent protein. The authors then wounded this dual-coloured 'fluorescent sandwich' and transplanted it into a healthy mouse. Comparison of the percentages of green and red cells revealed that 80% of cells in the healing wound came from the fascia. Furthermore, the vast majority of many cell types found in the healing injury originated from the fascia, including

contractile fibroblasts (or myofibroblasts), blood-vessel cells, macrophages of the immune system and nerve cells.

To confirm that their observations were not due to any peculiarities of this artificial grafted structure, the authors injected a dye into the fascia of mice, and then gave the mice a deep wound that penetrated the animals' skin and fascia. The authors mapped the dye-labelled cells that populated the healing wound and the surrounding scar tissue. More than half of the cells in the healed wound were labelled with the dye, confirming that the fascia is a major source of scar-forming tissue after deep injury.

Deep wounds lead to scars that are larger and harder to heal than those arising from superficial wounds that do not penetrate the fascia³. The authors used two-photon microscopy to analyse deep skin wounds in mice engineered⁴ to express fluorescent proteins, which can be used to trace scar-forming EPFs. They found that a cellular plug in the fascia, consisting of extracellular matrix, macrophages, blood vessels and nerves, moved upwards into the damaged skin to form a scar. This healing process did not require cell division, indicating that the plug was prefabricated. Importantly, the authors found that key proteins that have been reported to define the types of fibroblast found in scars⁵ are expressed at higher levels on fascial than on dermal fibroblasts, consistent with a model in which fascial EPFs are a major source of fibroblasts in healing deep wounds (Fig. 1).

Given that fibroblasts regulate the extracellular matrix, the authors used microscopy to visualize physical features of fibres of the protein collagen, which is a component of the extracellular matrix. Collagen in the fascia was more coiled and immature than were the stretched and interwoven collagen fibres in the dermis. Furthermore, when a fluorescent dye was used to tag collagen in an injured animal, this revealed that the extracellular matrix of the fascia moved upwards like a pliable gel into the damaged tissue, to

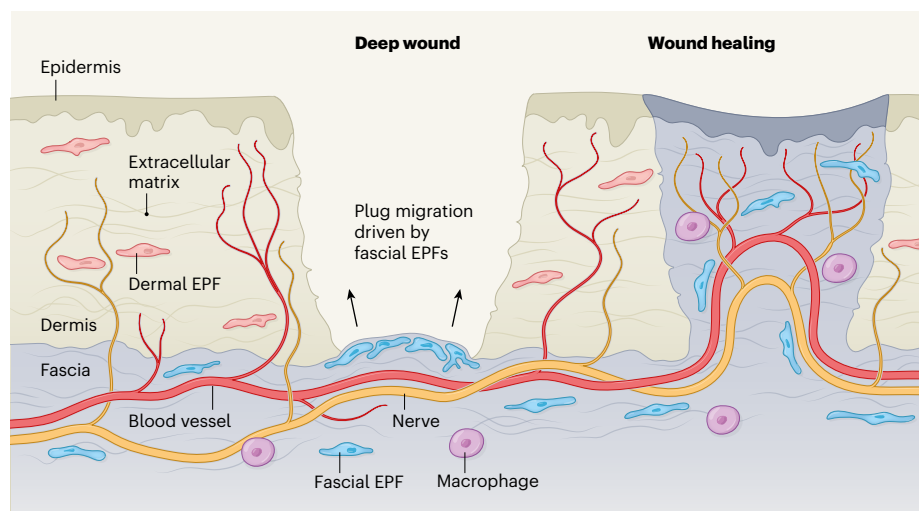


Figure 1 | The healing of deep skin wounds. The skin consists of an outer layer called the epidermis and an inner layer, the dermis. Superficial wounds no deeper than skin level can be repaired by cells called Engrailed-positive fibroblasts (EPFs) in the dermis, which make extracellular-matrix material. Working with mice, Correa-Gallegos *et al.*¹ investigated the healing of deep wounds that penetrated below the skin into a layer known as the fascia. The fascia contains EPFs, extracellular matrix, blood vessels, nerves and immune cells called macrophages. The authors report that a prefabricated plug of material from the fascia moves upwards, steered by fascial EPFs, to seal the wound. (Image based on Fig. 6 of ref. 1.)

plug and then repair the wound. By contrast, dermal collagen remained immobile.

The authors then tested whether EPFs from the fascia drive the movement of the prefabricated plug. They inserted non-adhesive membranes in mice to separate the fascia from the dermis, which resulted in delayed repair and non-healing wounds that remained open. Animals in which these membranes were not inserted did not show these effects. The removal of fascial EPFs by a genetic approach also resulted in the plug not entering wounds and in poor healing. These findings indicate that fascial EPFs do indeed steer the plug that seals deep wounds.

Although this study has potential relevance for human disease, most of the work was carried out in an artificial mouse model. Moreover, mice have a type of muscle called the panniculus carnosus, which lies between the fascia and the skin and is used to twitch the skin⁶. However, humans lack this twitching ability and have only a small remnant of this muscle. Therefore, the authors needed to determine whether scar formation occurs in a similar manner in humans and mice despite such differences.

The team analysed fascial fibroblasts in human skin and investigated a type of human raised scar called a keloid, which grows bigger than the original injury and can be profoundly itchy, inflamed and painful⁷. Many of the proteins that characterize the mouse fascia were also highly expressed in human fascia and keloid scars. This similarity suggests that the same processes are involved in wound healing and scar formation in both species. However, it is not yet clear whether these findings in mice reveal general principles that are relevant to human skin disease.

The authors' findings provide satisfying potential explanations for some unsolved clinical conundrums. Nerves, blood vessels and macrophages in the prefabricated plug are dragged into the mouse wound; if the same phenomenon occurs in humans, this could explain why keloids itch and are painful. Keloid formation is more common at sites of thicker fasciae (such as the chest, back and thighs) than at sites where the fascia is thinner (for example, the feet), which is consistent with a model in which the fascia drives keloid formation.

Could these discoveries about the skin shed light on other clinically relevant fibrotic diseases (conditions associated with the accumulation of extracellular matrix) that affect organs in which the fascia is not present, such as the lungs and liver? Perhaps the mechanisms uncovered in mice might have relevance for the processes underlying skin damage in the leg ulcers that can develop in people who have diabetes. In any case, it is clear that advances made in understanding the biology of the fascia

might reveal new targets for treating scarring diseases of the skin.

Mark C. Coles and **Christopher D. Buckley** are at the Kennedy Institute of Rheumatology, University of Oxford, Oxford OX3 7FY, UK.

C.D.B. is also at the Institute for Inflammation and Ageing, University of Birmingham, Birmingham, UK.

e-mails: c.d.buckley@bham.ac.uk;

mark.coles@kennedy.ox.ac.uk

1. Correa-Gallegos, D. *et al. Nature* **576**, 287–292 (2019).
2. Rinkevich, Y. *et al. Science* **17**, aaa2151 (2015).
3. Dunkin, C. S. *et al. Plast. Reconstr. Surg.* **119**, 1722–1732 (2007).
4. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. *Genesis* **45**, 593–605 (2007).
5. Driskell, R. R. & Watt, F. M. *Trends Cell Biol.* **25**, 92–99 (2014).
6. Stecco, C., Adstrum, S., Hedley, G., Schleip, R. & Yucesoy, C. A. *J. Bodyw. Mov. Ther.* **22**, 354 (2018).
7. Peng, G. L. & Kerolus, J. L. *Facial Plast. Surg. Clin. N. Am.* **27**, 513–517 (2019).

This article was published online on 27 November 2019.

Condensed-matter physics

Heat transferred in a previously unknown way

Karthik Sasiithlu

Experiments show that quantum fluctuations can allow heat to be transported between two objects separated by a vacuum gap. This effect could be harnessed to exploit and control heat transfer in nanoscale devices. **See p.243**

Acoustic waves and electromagnetic waves can transport heat between objects through their respective energy carriers: phonons and photons. At or near room temperature, the heat transfer between objects separated by a material medium occurs at a much higher

rate when facilitated by phonons than by photons. However, phonons are generally thought to be ineffective at transporting heat between objects separated by a vacuum gap, because these energy carriers are vibrations in an atomic lattice and thus would require a

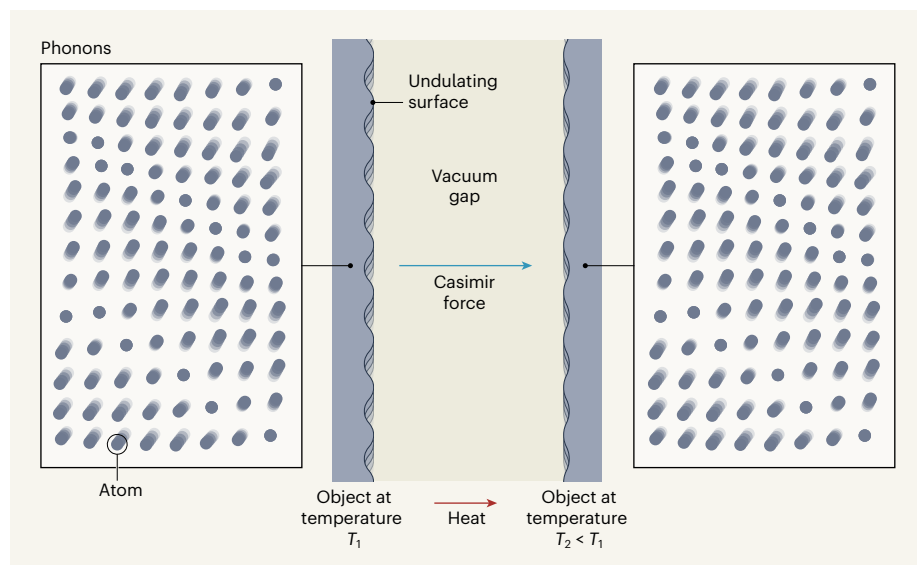


Figure 1 | Phonon transmission across a vacuum. Fong *et al.*¹ show that phonons – vibrations in an atomic lattice – can be transported between objects that are separated by a vacuum gap. To understand how this process occurs, consider an object at a fixed temperature T_1 . Thermal agitation of the object's atoms produces phonons that propagate as acoustic waves and cause the object's surface to exhibit time-varying undulations (the amplitudes of the undulations shown are exaggerated for clarity). A second object, at a fixed temperature $T_2 < T_1$, is brought close to the first object, with a vacuum gap between the objects. The undulations of the first object's surface exert a time-varying 'Casimir' force (caused by quantum fluctuations) on the second object's surface, which gives rise to phonons in the second object. Because phonons are heat carriers, heat is transferred from the first object to the second one.

material medium to propagate. On page 243, Fong *et al.*¹ report experimental evidence that phonons can travel across a vacuum gap and therefore induce heat transfer between vacuum-separated objects because of the effect of quantum fluctuations.

In simple terms, quantum fluctuations can be understood as being the source of an electromagnetic signal that a perfectly sensitive detector would detect in a vacuum, even when this vacuum is shielded from all possible internal and external sources of electromagnetic waves, such as charges and currents². The fluctuations are a consequence of a law in quantum mechanics known as Heisenberg's uncertainty principle³, which states that certain pairs of physical quantities cannot be determined at the same time with absolute precision. The presence of quantum fluctuations subtly influences surrounding matter, leading to several observable effects.

One of these effects, relevant to Fong and colleagues' work, is the Casimir force⁴ – the force that two neutral atoms separated by a vacuum gap exert on each other. The Casimir force results when quantum fluctuations induce fluctuating charge densities in these atoms; the charge densities then interact through their electric fields. The force that sticks a gecko's foot to a wall is an example of a macroscopic manifestation of the Casimir force. It arises from the combined interactions between fluctuating charge densities in all the atomic constituents of the two objects.

To understand how the Casimir force can induce phonon transfer between vacuum-separated objects, consider an object that is maintained at a particular temperature by being kept in contact with a heat source (Fig. 1). Thermal agitation of the object's atoms, which can be thought of as being interconnected by elastic springs, gives rise to phonons. In the presence of these phonons, the surface of the object undulates over time. When a second object is brought close to the first one, it is subjected to a time-varying Casimir force owing to its interaction with the undulations of the first object's surface. The second object's surface is thus subjected to tugging that then gives rise to phonons in the object's interior. Phonons are therefore transmitted from the first object to the second one.

Because phonons are heat carriers, when they are transported from one object to another across a vacuum gap, as a result of the Casimir force, they induce heat transfer if the second object is maintained at a lower temperature than that of the first one. This phenomenon of heat transport facilitated by the Casimir force has been predicted previously using theoretical models^{5–7}. Fong *et al.* have now measured such a heat-transfer mode experimentally.

The authors used a technique called optical interferometry to observe the thermal agitation of atoms (Brownian motion) at the surface

of a membrane. This membrane was kept in contact with a heat source held at a constant temperature. Measurements of thermal agitation can be related to, and therefore used as a gauge for, the temperature of the atoms at the membrane's surface. Moreover, the difference in this temperature with and without Casimir interaction with another, closely juxtaposed membrane is directly proportional

“Fong and colleagues' work provides conclusive evidence that the Casimir force can induce heat transfer.”

to the resulting heat transfer between the two interacting membranes. The authors used these features to estimate the amount of heat transmitted between the membranes for vacuum gaps of different sizes. They found that their measurements accurately conform to theoretical estimates of such heat transport.

Fong and colleagues' work provides conclusive evidence that the Casimir force can induce heat transfer. However, the use of this method to transport heat between two objects is limited, because the Casimir force decreases rapidly in strength as the space between the objects is increased. It is only when the gap between two objects is of the order of a few nanometres that the Casimir force is strong enough for this heat-transfer mode to dominate over

competing modes, such as photon tunnelling⁸.

The authors discovered a way to amplify the Casimir mode of heat transfer so that it remains dominant even when the gap between the membranes is in the range of hundreds of nanometres. The membranes were carefully designed in such a way that their dimensions and the temperatures at which they were maintained allowed them to vibrate with their maximum possible displacements – in other words, at their natural frequencies. Thus, applications that are devised to exploit this heat-transfer mode to dissipate heat (such as in a hard-disk drive, where the distance between the writing head and the storage disk is a few nanometres) would require such careful design to ensure that the mode is amplified. Achieving this would be a challenge for the future.

Karthik Sasihithlu is in the Department of Energy Science and Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India.
e-mail: ksasihithlu@iitb.ac.in

1. Fong, K. Y. *et al.* *Nature* **576**, 243–247 (2019).
2. Simpson, W. M. R. & Leonhardt, U. in *Forces of the Quantum Vacuum: An Introduction to Casimir Physics* 26 (World Scientific, 2015).
3. Heisenberg, W. *Z. Phys.* **43**, 172–198 (1927).
4. Casimir, H. B. G. *Proc. K. Ned. Akad. Wet. B* **51**, 793–795 (1948).
5. Budaev, B. V. & Bogoy, D. B. *Appl. Phys. Lett.* **99**, 053109 (2011).
6. Ezzahri, Y. & Joulain, K. *Phys. Rev. B* **90**, 115433 (2014).
7. Pendry, J. B., Sasihithlu, K. & Craster, R. V. *Phys. Rev. B* **94**, 075414 (2016).
8. Kim, K. *et al.* *Nature* **528**, 387–391 (2015).

Structural biology

Malaria parasites fine-tune mutations to resist drugs

Leann Tilley & Philip J. Rosenthal

Drug resistance in malaria parasites is mediated by mutations in a transporter protein. The transporter's structure reveals the molecular basis of how key mutations bring about resistance to different drugs. **See p.315**

About half a million people, most of them children living in Africa, are killed each year by malaria¹. Management of malaria, particularly that caused by the highly virulent protozoan parasite *Plasmodium falciparum*, is challenged by the emergence of resistance to antimalarial drugs². On page 315, Kim *et al.*³ report the structure and molecular properties of a key protein that facilitates resistance, the *P. falciparum* chloroquine-resistance transporter (PfCRT). The structure reveals the consequences of finely tuned mutations of the amino-acid residues that line a crucial central

cavity in PfCRT. These mutated residues allow resistant parasites to transport certain antimalarial drugs away from their site of action – and the effect of the mutations is different for closely related drugs.

Malaria parasites spend part of their life cycle inside human red blood cells. There, they use a specialized membrane-bound compartment known as the digestive vacuole to degrade the protein haemoglobin, thereby generating amino-acid building blocks for growth⁴. Haemoglobin digestion also produces a toxic side product called haem,

News & views

which is exploited by antimalarials from the quinoline family of drugs (which includes chloroquine and piperazine). These drugs bind to the released haem in the digestive vacuole and prevent the compound's detoxification by the parasite – in effect, poisoning the parasite with its own metabolic debris⁴.

Chloroquine's affordability, safety and efficacy made it the drug of choice for combating malaria, until widespread resistance developed in the 1980s. Piperazine is a structurally related compound that retains activity against chloroquine-resistant parasites and is currently used in combination with another antimalarial drug, dihydroartemisinin. Unfortunately, resistance to piperazine is now widespread in parts of southeast Asia⁵. Paradoxically, piperazine-resistant parasites are often more sensitive to chloroquine than are piperazine-sensitive parasites⁶.

Drug resistance can arise because of mutations either in the drug's target or in biological machinery that transports the drug to or from the target. A landmark report⁷ in 2000 identified PfCRT as the main mediator of chloroquine resistance in *P. falciparum*. Resistance emerged independently at different locations around the world, but is always associated with a particular mutation in the transporter – the substitution of a lysine amino-acid residue by a threonine residue (a K76T mutation). K76T combines with other geographically specific PfCRT mutations to mediate resistance and improve the fitness of the mutant parasites^{6–8}.

PfCRT is a member of the superfamily of proteins known as drug/metabolite transporters⁹ and is located in the membrane of the parasite's digestive vacuole. Structural and biochemical studies of related proteins suggest that resistance to chloroquine arises as a result of the transporter passing chloroquine out of the digestive vacuole, thus removing it from its site of action⁸. Interestingly, piperazine resistance emerges when parasites harbouring the K76T mutation acquire further mutations⁶.

Efforts to design new resistance-busting quinoline antimalarials have been stymied by a lack of information about the structures of mutated PfCRT. But in the past few years, advances in a technique called cryo-electron microscopy¹⁰ (cryo-EM) have revolutionized structural biology by enabling the direct imaging of membrane-embedded proteins, such as PfCRT, that are not amenable to study using X-ray crystallography. Kim *et al.* used single-particle cryo-EM to determine the structure of PfCRT in South American 7G8 parasites, which harbour mutations that confer high-level chloroquine resistance.

The authors first had to work out a protocol for efficiently expressing, purifying and reconstituting PfCRT into a membrane-like environment, to maintain a native conformation of the protein. PfCRT is relatively small

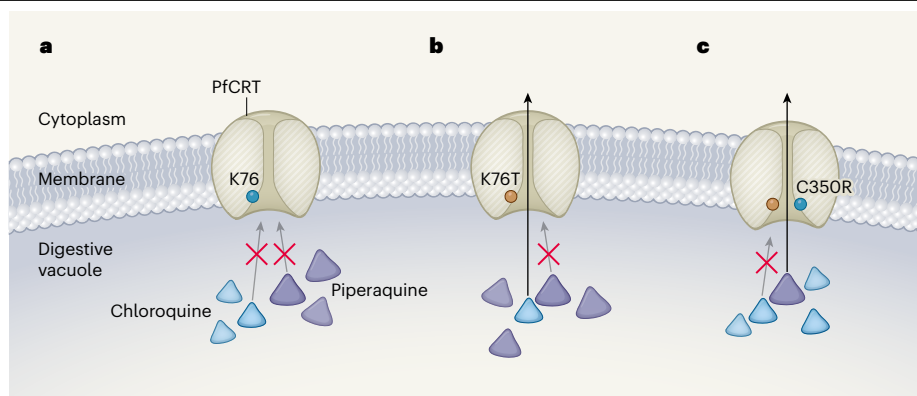


Figure 1 | Mutations affect the transport of antimalarial drugs through the PfCRT protein. **a**, The drugs chloroquine and piperazine target a membrane-bound organelle in the malaria parasite, called the digestive vacuole. *Plasmodium falciparum* chloroquine-resistance transporter (PfCRT) resides in the membrane of the digestive vacuole of the most virulent species of malaria parasite. A positively charged lysine amino-acid residue (K76) in wild-type PfCRT is thought to repel the positively charged drugs, preventing transport out of the vacuole (arrow) – which would lead to drug resistance. **b**, Kim *et al.*³ report the structure of PfCRT that contains a mutation known as K76T. The authors find that K76T alters the charge distribution of the lining of the PfCRT cavity. Both chloroquine and piperazine bind to the mutated cavity, but only chloroquine passes through it (possibly because it has a lower positive charge than piperazine), enabling resistance to chloroquine, but not to piperazine. **c**, Further mutations in PfCRT (such as the C350R mutation) cause piperazine to bind more weakly to the cavity than in **b**. This potentially underpins the drug's ability to pass through PfCRT and leads to piperazine resistance. In this scenario, the transport of chloroquine out of the digestive vacuole is less efficient than the transport of piperazine.

(49 kilodaltons) compared with proteins that are typically studied using cryo-EM, and so Kim *et al.* prepared an antibody fragment (known as an antigen-binding fragment, or Fab) that binds to PfCRT, thereby forming a complex that has sufficient mass and stability to allow cryo-EM-based structural elucidation. This approach yielded a structure with 3.2-ångström resolution.

PfCRT was revealed to have ten transmembrane domains and a negatively charged central cavity. The cavity opens on to the digestive vacuole, but closes about halfway through the membrane. The Fab-bound cavity has an opening 25 Å in diameter, which is large enough to contain chloroquine (the maximum dimension of which is about 14 Å) and piperazine (with a maximum dimension of about 21 Å). The structure shows that wild-type PfCRT has a positively charged lysine residue (K76) positioned in the cavity. This residue is thought to repel both chloroquine (which has two positive charges) and piperazine (which has four positive charges), thereby trapping them in the digestive vacuole (Fig. 1a).

Interestingly, Kim and colleagues' biochemical experiments show that PfCRT from 7G8 parasites can bind both chloroquine and piperazine, but transport only chloroquine; thus, parasites with the K76T mutation are resistant to chloroquine, but not piperazine (Fig. 1b). By contrast, when the authors introduced further mutations to 7G8 that have been observed in South American¹¹ and southeast Asian⁶ parasites in the past five years, piperazine efflux increased – an effect that was associated with decreased sensitivity to piperazine, but

increased sensitivity to chloroquine (Fig. 1c).

Kim *et al.* used their cryo-EM structure to carry out molecular modelling and electrostatic analysis of PfCRT, to help explain why mutations can have opposite effects on the sensitivity of malaria parasites to chloroquine and piperazine. The modelling suggests that mutations associated with piperazine resistance can reduce the negative charge or alter the conformation of the PfCRT central cavity. This might prevent piperazine from binding too tightly to the cavity and thus increase its transport out of the digestive vacuole. The authors propose that the distribution of surface charges in the cavity can be fine-tuned so that the initial binding of a drug to PfCRT, and its subsequent release for transport, is different for different drugs, thereby producing distinct effects on drug sensitivity. Taken together, Kim and colleagues' findings show that *P. falciparum* is engaged in an ongoing balancing act, generating mutations that block the action of different drugs while maintaining optimal fitness of the parasite.

A limitation of the new structure is that PfCRT is locked in a conformation in which it is open to the digestive vacuole, as a result of Fab binding in its central cavity. Further studies will be required to work out how drug binding couples to the conformational rearrangements that permit drug transport.

On a practical level, it is particularly important to understand the mechanisms of resistance that are likely to arise in Africa, where more than 90% of cases of malaria caused by *P. falciparum* occur¹. Compared with other parts of the world, there are fewer PfCRT mutations

in Africa, where resistance to chloroquine is decreasing in many countries and where the combination of piperazine and dihydroartemisinin remains highly effective¹². We can look forward to further studies of PfCRT, including visualization of the drug-bound and open-to-cytoplasm structural conformations, which will further explain the effects of resistance mutations and might help to identify drugs that circumvent resistance. For now, we can appreciate the insights gained from Kim and colleagues' beautiful marriage of structure, biochemistry, genetics and parasitology, and particularly from the first atomic-resolution structure of PfCRT — the fine-tuned, resistance-mediating machine of malaria parasites.

Leann Tilley is in the Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia. **Philip J. Rosenthal** is in

the Department of Medicine, University of California, San Francisco, San Francisco, California 94143 USA.
e-mails: ltilley@unimelb.edu.au;
philip.rosenthal@ucsf.edu

1. World Health Organization. *World Malaria Report 2018* (2018).
2. Haldar, K., Bhattacharjee, S. & Safeukui, I. *Nature Rev. Microbiol.* **16**, 156–170 (2018).
3. Kim, L. *et al. Nature* **576**, 315–320 (2019).
4. Goldberg, D. E. *Curr. Top. Microbiol. Immunol.* **295**, 275–291 (2005).
5. van der Pluijm, R. W. *et al. Lancet Infect. Dis.* **19**, 952–961 (2019).
6. Ross, L. S. *et al. Nature Commun.* **9**, 3314 (2018).
7. Fidock, D. A. *et al. Mol. Cell* **6**, 861–871 (2000).
8. Summers, R. L. *et al. Proc. Natl Acad. Sci. USA* **111**, E1759–E1767 (2014).
9. Martin, R. E. & Kirk, K. *Mol. Biol. Evol.* **21**, 1938–1949 (2004).
10. Nogales, E. & Scheres, S. H. *Mol. Cell* **58**, 677–689 (2015).
11. Pelleau, S. *et al. Proc. Natl Acad. Sci. USA* **112**, 11672–11677 (2015).
12. Conrad, M. D. & Rosenthal, P. J. *Lancet Infect. Dis.* **19**, e338–e351 (2019).

This article was published online on 27 November 2019.

Solar physics

A step closer to the Sun's secrets

Daniel Verscharen

NASA's Parker Solar Probe is currently making a series of close encounters with the Sun. Initial observations from the spacecraft have improved our understanding of both the Sun and its environment. **See p.223, p.228, p.232 & p.237**

Although the Sun is quite near to us compared with other stars, it has always kept intriguing and fundamental scientific secrets from us. For instance, we still don't know how the solar corona — the Sun's outermost atmosphere — maintains temperatures in excess of one million kelvin, whereas the visible surface has temperatures of just below 6,000 K (ref. 1). The corona produces the solar wind, an outflow of plasma particles (free ions and electrons) that expands into the space between the planets. In 2018, NASA launched the Parker Solar Probe² (PSP) with the aim of identifying the mechanisms behind the heating of the corona and the acceleration of the solar wind. Four papers in *Nature*^{3–6} report the first results from the PSP.

The measurements from the PSP were taken when the spacecraft was as close as 24 million kilometres to the Sun (for comparison, the average distance between Mercury and the Sun is about 58 million kilometres). They show that the solar wind near the Sun is much more structured and dynamic than it is at Earth (Fig. 1). On page 237, Bale *et al.*³ present measurements of the direction and

strength of the Sun's magnetic field, which is dragged out into space by the solar wind. The authors find rapid reversals in the direction of the field that last for only minutes.

Although some similar magnetic structures have been seen before⁷, the large amplitude and the high occurrence rate of these reversals are surprising. In fact, the nature of these structures remains unknown.

Bale and colleagues also report that the PSP's sensors detected fluctuations in the local electric and magnetic fields in the solar wind that are larger than those detected near Earth. These fluctuations can be generated by turbulence in the solar wind or by plasma instabilities that are driven by ions or electrons. The presence of such fluctuations suggests that plasma instabilities have a much larger effect on the dynamics and energetics of the solar wind than previously expected.

On page 228, Kasper *et al.*⁴ present observations of the Sun's plasma ions and electrons. They find that the reversals in the Sun's magnetic field are often associated with localized enhancements in the radial component of the plasma velocity (the velocity in the direction away from the Sun's centre). The authors use the extremely clear signal of the solar wind's *strahl* — a collimated and fast beam of electrons that stream along the magnetic field — to study the field's geometry and configuration. This method leads Kasper and colleagues to interpret the magnetic-field reversals as travelling S-shaped bends in the field lines coming from the Sun.

These authors also report a surprisingly large azimuthal component of the plasma velocity (the velocity perpendicular to the radial direction). This component results from the force with which the Sun's rotation slingshots plasma out of the corona when the plasma is released from the coronal magnetic field — much like a spinning hammer-thrower slingshots the hammer when releasing it from their hands. However, the reason for the large observed value of the azimuthal velocity is currently unclear.

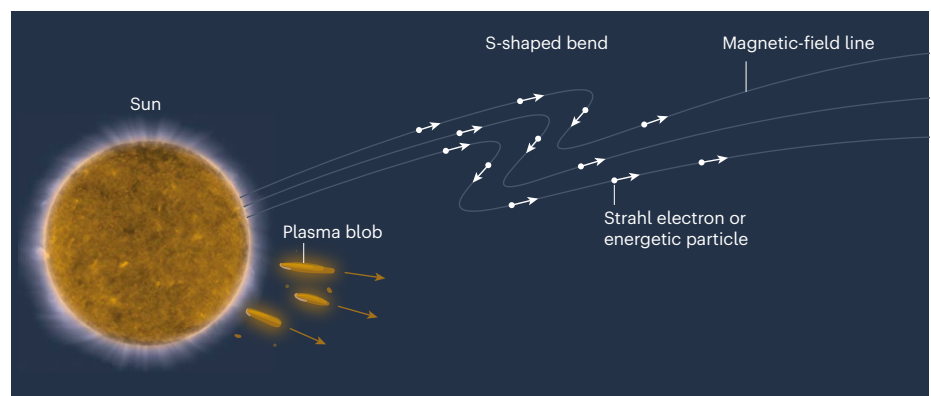


Figure 1 | The near-Sun environment. The Sun's outermost atmosphere generates an outflow of plasma particles (ions and electrons) called the solar wind. 'Strahl' electrons and energetic particles in the wind stream along the Sun's magnetic-field lines. Four papers^{3–6} report observations from the Parker Solar Probe (PSP), which is currently in orbit around the Sun. The PSP data suggest that the field lines contain S-shaped bends and that the Sun releases blobs of plasma that form part of the young solar wind. The ultraviolet-light image of the Sun was taken by NASA's Solar Dynamics Observatory on the day that the PSP made its first close encounter with the Sun.

On page 223, McComas *et al.*⁵ study detections of energetic ions and electrons, some of which are observed more often in the region just outside the corona than they are near Earth. These particles are accelerated by flares (eruptions of radiation) in the corona or by shock waves associated with coronal mass ejections (eruptions of plasma), which travel through interplanetary space. The authors identify particles corresponding to both types of source region.

Because energetic particles travel along the Sun's magnetic field, the difference in the time at which fast and slow particles arrive at the PSP can be used to estimate the path length of their trajectory along the field. McComas and colleagues find that this path length is longer than expected, which suggests that the magnetic field has a more complicated geometry than assumed. This finding could be accounted for by the S-shaped magnetic-field reversals.

The imaging instrument on board the PSP makes remote observations of light scattered by electrons and dust near the Sun. On page 232, Howard *et al.*⁶ report that the intensity of the dust-scattered light decreases with distance from the Sun in almost the same way as it does when observed from Earth. However, the authors find some preliminary evidence for the existence of a hypothesized dust-free zone⁸ near the Sun that has not been detected before. The detailed images from the PSP also show spatial variations in the solar wind that are consistent with variations in the Sun's magnetic field on its surface, and reveal small blobs of plasma that are ejected from the Sun and form part of the young solar wind.

These four papers show that, by going into an unexplored region of the Solar System, the PSP has already made great discoveries. In the near future, it will be important to combine all the available sources of information to develop a deeper understanding of the physics of the Sun and the solar wind. For instance, researchers should combine the measurements of the electric and magnetic fields with detailed observations of the plasma particles to determine how fields and plasma interact and drive instabilities⁹. They must also study the large azimuthal flow velocity further to confirm whether it is a persistent feature or just a one-time exception during these initial PSP measurements.

The use of magnetic-field models will enable scientists to learn more about the path of energetic particles between the Sun and the PSP, and, in turn, about space weather – the effects of the Sun and the solar wind on Earth and human technology. These energetic-particle studies must also be linked with remote observations of the Sun's surface and the corona. Examining the potential presence of the dust-free zone near the Sun must be another short-term goal, but might have to

wait for closer approaches of the PSP to the Sun in the future.

It is expected that PSP data will guide our understanding of the Sun and the solar wind for many years. New models and theories will be motivated by the spacecraft's discoveries, and this knowledge will be transferable to other stars and astrophysical plasmas throughout the Universe. After all, the Sun is the only star that we can study up close using spacecraft. The orbit of the PSP will bring the spacecraft even closer to the Sun in the coming years, to just over 6 million kilometres from the surface². During this time, the Sun will transition into a more active phase of its 11-year cycle, so we can expect even more-exciting results soon.

In 2020, the European Space Agency will launch the Solar Orbiter mission¹⁰. Although this spacecraft will not go quite as close to the Sun as will the PSP, its more extensive suite of scientific instruments will be used in combination with the PSP to reveal key information about the Sun. For example, Solar Orbiter will measure the elemental

composition and charge states of ions and will take photographs of the Sun in different wavelengths of light. These joint measurements will certainly close some of the remaining gaps in our knowledge of the Sun and the solar wind. For now, however, the Sun has proved again that it still holds more secrets for us to discover.

Daniel Verscharen is at the Mullard Space Science Laboratory, University College London, Dorking RH5 6NT, UK.
e-mail: d.verscharen@ucl.ac.uk

1. Aschwanden, M. *Physics of the Solar Corona: An Introduction with Problems and Solutions* (Springer, 2005).
2. Fox, N. J. *et al. Space Sci. Rev.* **204**, 7–48 (2016).
3. Bale, S. D. *et al. Nature* **576**, 237–242 (2019).
4. Kasper, J. C. *et al. Nature* **576**, 228–231 (2019).
5. McComas, D. J. *et al. Nature* **576**, 223–227 (2019).
6. Howard, R. A. *et al. Nature* **576**, 232–236 (2019).
7. Horbury, T. S. *et al. Mon. Not. R. Astron. Soc.* **478**, 1980–1986 (2018).
8. Lamy, P. L. *Astron. Astrophys.* **33**, 191–194 (1974).
9. Marsch, E. *Living Rev. Sol. Phys.* **3**, 1 (2006).
10. Müller, D., Marsden, R. G., St. Cyr, O. C. & Gilbert, H. R. *Solar Phys.* **285**, 25–70 (2013).

This article was published online on 4 December 2019.

Electronics

Graphene sees the light in wearable sensors

Deji Akinwande & Dmitry Kireev

Graphene coated with nanoparticles has been used to make wearable light sensors that measure the human pulse and blood oxygen levels from ambient light passing through tissue, offering a potential platform for health-care monitoring.

The popularity of wearable technology has risen enormously, with the US market projected to be in the tens of billions of dollars by 2022 (see go.nature.com/33tcein). However, the effectiveness of the most common wearable devices is hindered by the physical specifications of their components: although the device is often embedded in a flexible soft shell, the main parts, such as the sensors and electronics, are still rigid^{1,2}. Now, writing in *Science Advances*, Polat *et al.*³ report a class of truly flexible, transparent wearable device that is based on graphene covered with a layer of semiconducting nanoparticles known as quantum dots. Impressively, the devices measure various vital signs using only ambient light as a signal.

Materials that are just one or a few atoms thick are said to be two-dimensional. The best-known example is graphene, which consists of single sheets of carbon atoms arranged in a hexagonal lattice. 2D materials

in general, and graphene in particular, have tremendous potential for the development of next-generation wearable, soft biosensors because they combine electrical conductivity, optical transparency and mechanical flexibility with outstanding biocompatibility⁴ and stability to biological electrolytes. Graphene-based tattoo-like devices⁵ have previously been used to record human health signals such as heart rhythm, skin hydration and body temperature. Their outstanding performance is associated with the subnanometre thickness of graphene, which allows it to bend and stretch with the skin, without affecting the sensor performance.

Polat *et al.* have now expanded the functionality of graphene in wearable devices by depositing light-sensitive quantum dots made of the semiconductor lead(II) sulfide (PbS) onto the graphene layer. When illuminated, the quantum dots generate pairs of charged particles: negatively charged electrons and

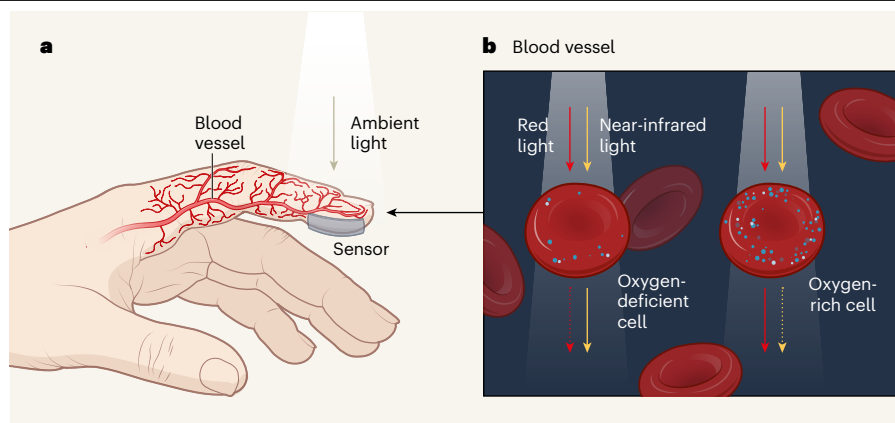


Figure 1 | A sensor that detects vital signs using light. **a**, Polat *et al.*³ have used graphene coated with semiconducting nanoparticles (not shown) to make flexible, transparent devices that can detect the light transmitted through tissue. Ambient light passes easily through human skin and tissue, but is absorbed by the haemoglobin in red blood cells. By monitoring the modulation of ambient light as it passes through tissue, the sensors can thus monitor blood flow and heart rate. The rate of respiration can also be estimated from mathematical analysis of the blood-flow data. **b**, Moreover, absorption of red and near-infrared light by oxygen-rich red blood cells is significantly different from that by oxygen-poor cells. By monitoring the absorption of red and near-infrared light, the sensors can monitor the oxygen content of the blood.

positively charged holes (quasiparticles associated with the absence of an electron in an atomic lattice). The electrons stay trapped in the quantum dots, but the holes are transferred into the graphene layer and increase its electrical conductivity, producing a measurable electrical signal. The authors used this behaviour to construct light sensors from the quantum-dot-coated graphene.

The researchers observed that the responsivity (the electrical output per optical input) of their devices was remarkably large. The high responsivity is attributable to the fact that the holes in the graphene layer are recycled by the quantum dots, effectively increasing the number of charge carriers generated per absorbed photon in the devices – the devices are said to exhibit a photoconductive gain.

Previously reported light sensors typically do not have photoconductive gain, and therefore require an amplifier device to boost the electrical signal; this increases both power consumption and the size of the overall device⁶. Moreover, the amplifier must be in close proximity to the sensor, which can limit the ability of wearable devices to take on the contour of the skin. The intrinsic photoconductive gain of Polat and colleagues' devices eliminates the need for an amplifier, solving the above problems and making the sensors particularly suitable for real-life applications.

So how were the sensors used to measure vital signs? Light at certain wavelengths passes easily through human skin and adjacent tissue, but is absorbed strongly by blood⁷ – more specifically, it is absorbed by haemoglobin, the molecule that transports oxygen in red blood cells. By continuously monitoring the intensity of light passing

through tissue, sensors can produce readouts called photoplethysmograms (PPGs) that contain information about volumetric changes to blood vessels, which can be correlated to heart rate⁸. Polat and colleagues show that their wearable devices can, remarkably, use the ambient light that passes through tissue to measure human heart rates accurately. Moreover, the sensitivity of the devices allowed the researchers to estimate the rate of breathing by mathematically analysing the PPG data. Physical movements associated with breathing usually produce artefacts and noise in the PPG signals detected by rigid wearable devices⁹, but the physical unobtrusiveness and flexibility of the new devices overcome this problem.

“Graphene has paved the way for other 2D materials to be used in sensors and mobile health-monitoring devices.”

Polat *et al.* report that their wearable devices can also monitor another vital health signal that is often checked by doctors: arterial oxygen saturation (SpO₂), which is the percentage of haemoglobin in blood that is loaded with oxygen (Fig. 1). Low SpO₂ levels can result in loss of consciousness, impaired mental functions, and respiratory and cardiac arrest. The absorption of red light and near-infrared light by oxygen-rich red blood cells is significantly different from the absorption by oxygen-free cells. The authors therefore estimated SpO₂ levels by using their devices to measure light absorption at these two wavelengths.

Finally, Polat and colleagues' reported a

further application of their technology: the monitoring of ultraviolet light. Certain UV wavelengths can be harmful to the skin, and can potentially even cause cancer¹⁰, making it desirable to measure UV levels in the environment. The authors show that their devices can be integrated with previously fabricated chips that enable the sensors to wirelessly transfer UV measurements to a mobile phone, thus enabling continuous and convenient monitoring of the environmental UV index.

The reported sensors are all designed to communicate wirelessly to any other electronics needed for a wearable device, clearly separating the soft sensor from any rigid components. But the wireless design requires a read-out device (such as a mobile phone) to be close to the sensor, which makes it difficult to perform long-term monitoring – as might be needed for heart-rate monitoring, for example. Establishing long-term, continuous communication between the wearable flexible sensors and conventional electronics will be essential for future applications. Alternatively, it might be possible to include components that enable memory storage and simple digital processing in the flexible platform. This could be achieved in the future using 2D materials other than graphene¹¹.

Graphene has now been used as a sensor and as a signal transducer in various prototypes for wearable and mobile health devices¹². More importantly, however, graphene has paved the way for other 2D materials to be used in sensors and mobile health-monitoring devices. Thousands of such materials have been discovered, with as-yet unknown properties¹³. We think that the comprehensive study of those materials will be essential for the development of future biosensors that can be worn by, or even integrated into, humans.

Deji Akinwande and **Dmitry Kireev** are in the Department of Electrical and Computer Engineering, Microelectronics Research Center, The University of Texas at Austin, Austin, Texas 78758, USA.
e-mails: deji@ece.utexas.edu;
d.kireev@utexas.edu

- Huang, Z. *et al.* *Nature Electron.* **1**, 473–480 (2018).
- Kim, D.-H. *et al.* *Science* **333**, 838–843 (2011).
- Polat, E. O. *et al.* *Sci. Adv.* **5**, eaaw7846 (2019).
- Fabbro, A. *et al.* *ACS Nano* **10**, 615–623 (2016).
- Kabiri Ameri, S. *et al.* *ACS Nano* **11**, 7634–7641 (2017).
- Kim, J., Kim, J. & Ko, H. *Sensors* **16**, E46 (2015).
- Mannheimer, P. D. *Anesth. Analg.* **105**, S10–S17 (2007).
- Allen, J. *Physiol. Meas.* **28** (3), R1–R39 (2007).
- Shimazaki, T. & Hara, S. in *Proc. 2015 9th Int. Symp. Med. Inform. Commun. Technol.* 200–203 (IEEE, 2015).
- Kuluncsics, Z., Perdiz, D., Brulay, E., Muel, B. & Sage, E. *J. Photochem. Photobiol. B* **49**, 71–80 (1999).
- Akinwande, D. *et al.* *Nature* **573**, 507–518 (2019).
- Huang, H. *et al.* *Front. Chem.* **7**, 399 (2019).
- Ferrari, A. C. *et al.* *Nanoscale* **7**, 4598–4810 (2015).

This article was published online on 18 November 2019.

Probing the energetic particle environment near the Sun

<https://doi.org/10.1038/s41586-019-1811-1>

Received: 28 June 2019

Accepted: 5 September 2019

Published online: 4 December 2019

D. J. McComas^{1*}, E. R. Christian², C. M. S. Cohen³, A. C. Cummings³, A. J. Davis³, M. I. Desai^{4,5}, J. Giacalone⁶, M. E. Hill⁷, C. J. Joyce¹, S. M. Krimigis⁷, A. W. Labrador³, R. A. Leske³, O. Malandraki⁸, W. H. Matthaeus⁹, R. L. McNutt Jr⁷, R. A. Mewaldt³, D. G. Mitchell⁷, A. Posner¹⁰, J. S. Rankin¹, E. C. Roelof⁷, N. A. Schwadron^{1,11}, E. C. Stone³, J. R. Szalay¹, M. E. Wiedenbeck¹², S. D. Bale^{13,14}, J. C. Kasper¹⁵, A. W. Case¹⁶, K. E. Korreck¹⁶, R. J. MacDowall², M. Pulupa¹³, M. L. Stevens¹⁶ & A. P. Rouillard¹⁷

NASA's Parker Solar Probe mission¹ recently plunged through the inner heliosphere of the Sun to its perihelia, about 24 million kilometres from the Sun. Previous studies farther from the Sun (performed mostly at a distance of 1 astronomical unit) indicate that solar energetic particles are accelerated from a few kiloelectronvolts up to near-relativistic energies via at least two processes: 'impulsive' events, which are usually associated with magnetic reconnection in solar flares and are typically enriched in electrons, helium-3 and heavier ions², and 'gradual' events^{3,4}, which are typically associated with large coronal-mass-ejection-driven shocks and compressions moving through the corona and inner solar wind and are the dominant source of protons with energies between 1 and 10 megaelectronvolts. However, some events show aspects of both processes and the electron–proton ratio is not bimodally distributed, as would be expected if there were only two possible processes⁵. These processes have been very difficult to resolve from prior observations, owing to the various transport effects that affect the energetic particle population en route to more distant spacecraft⁶. Here we report observations of the near-Sun energetic particle radiation environment over the first two orbits of the probe. We find a variety of energetic particle events accelerated both locally and remotely including by corotating interaction regions, impulsive events driven by acceleration near the Sun, and an event related to a coronal mass ejection. We provide direct observations of the energetic particle radiation environment in the region just above the corona of the Sun and directly explore the physics of particle acceleration and transport.

Onboard the Parker Solar Probe (PSP), the instrument suite of the Integrated Science Investigation of the Sun (IS²IS)⁷ has made the first near-Sun measurements of solar energetic particles (SEPs). IS²IS comprises two energetic particle instruments with overlapping coverage, EPI-Hi and EPI-Lo, measuring higher- and lower-energy particles⁷. Together they enable IS²IS to explore the near-Sun environment by measuring the fluxes, energy spectra, anisotropy, and composition of suprathermal and energetic ions with energies from about 0.02 to 200 MeV per nucleon and electrons with energies from about 0.05 to 6 MeV. Here we examine this energetic particle environment in the context of in situ solar wind⁸ and magnetic field⁹ conditions and surrounding density structures¹⁰ measured by other instruments onboard PSP.

Figure 1 summarizes IS²IS observations of energetic particles over PSP's first two orbits. Higher- (1–2 MeV) and lower-energy (30–200 keV)

H⁺ ion count rates are plotted on the outside and inside of the orbital trajectory, respectively. Intensifications indicate energetic particle events, with some seen only at higher energies, some only at lower energies, and others simultaneously across the combined energy range. Figure 1 indicates how rich the IS²IS observations are: a broad array of different types of particle events are seen at all distances.

The first large intensification occurred during orbit 1 at higher energies with PSP inbound (during interval a, 2018-287 18:00 to 2018-297 08:20 universal time (UT)) at about 0.5 astronomical units (AU). Although not obvious from Fig. 1, this is a corotational event also seen when PSP was outbound at about 0.65 AU (during interval b, 2018-330 23:20 to 2018-341 15:00 UT). Corotating interaction regions (CIRs) form as faster solar wind piles up behind slower wind, forming a compression^{11,12}. Because these faster solar wind streams

¹Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. ²Goddard Space Flight Center, Greenbelt, MD, USA. ³California Institute of Technology, Pasadena, CA, USA.

⁴Southwest Research Institute, San Antonio, TX, USA. ⁵University of Texas at San Antonio, San Antonio, TX, USA. ⁶University of Arizona, Tucson, AZ, USA. ⁷Johns Hopkins University Applied

Physics Laboratory, Laurel, MD, USA. ⁸National Observatory of Athens, IAASARS, Athens, Greece. ⁹University of Delaware, Newark, DE, USA. ¹⁰NASA HQ, Washington, DC, USA. ¹¹University of

New Hampshire, Durham, NH, USA. ¹²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. ¹³University of California at Berkeley, Berkeley, CA, USA. ¹⁴The Blackett

Laboratory, Imperial College London, London, UK. ¹⁵University of Michigan, Ann Arbor, MI, USA. ¹⁶Smithsonian Astrophysical Observatory, Cambridge, MA, USA. ¹⁷CNRS, Toulouse, France.

*e-mail: dmccomas@princeton.edu

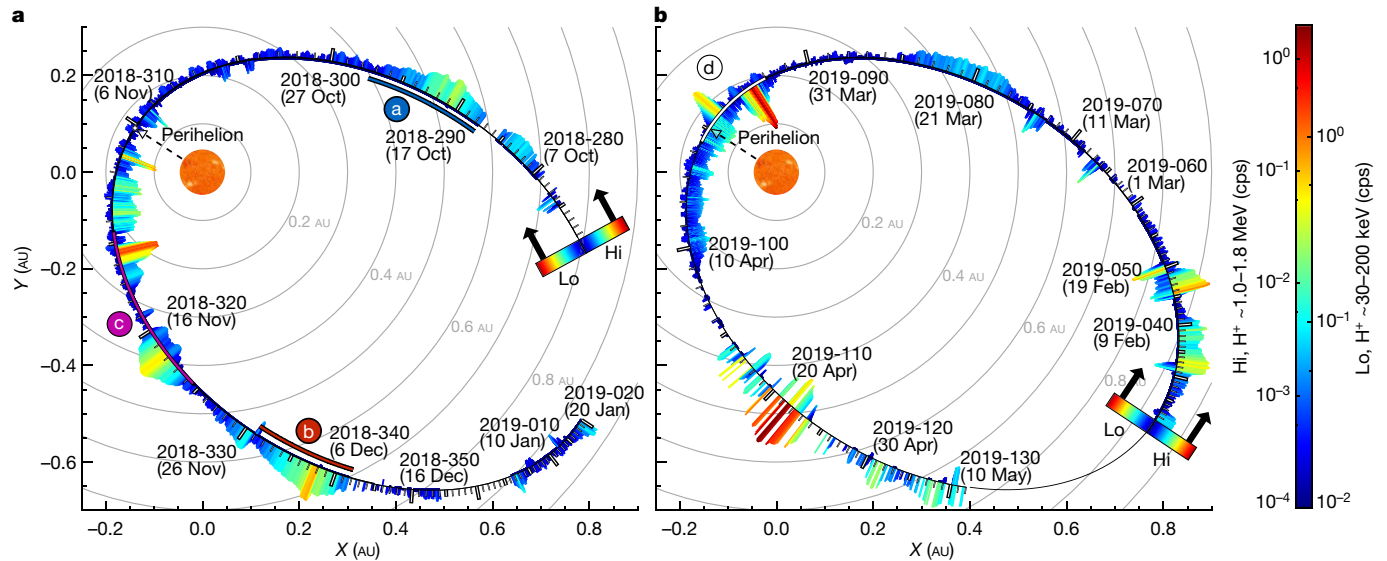


Fig. 1 | Observations of energetic particles during orbits 1 and 2. a, b, Observations of energetic particles (primarily H^+) at lower (Lo, about 30–200 keV; inside track) and higher energies (Hi, about 1–2 MeV; outside track) from PSP's orbit 1 (a) and orbit 2 (b). Intervals without data are indicated by the

black orbital track. Particle intensity is indicated by both the colour and the length of the bars. Intervals a–d are earmarked for detailed study. The scale indicated by white rectangles on the outer track is measured in days UT, from 2018-280 (7 October 2018) to 2019-020 (20 January 2019).

emanate from coronal holes at the Sun, CIRs map to nearly fixed solar longitudes.

Figure 2 shows intervals a and b as a function of the longitude of the solar surface ‘foot point’ magnetically connected to the spacecraft, calculated for a nominal Parker spiral with a fixed solar wind speed of 350 km s^{-1} . This calculation combines the rotation of the Sun and spacecraft location to show that both events arise from the same, single CIR

structure. These events are ‘dispersionless’—all ions arrive at roughly the same time and fluctuations in intensity are consistent across ion speeds. Such events indicate that PSP passed across magnetic flux tubes that were already filled with high-energy ($>1 \text{ MeV}$) particles that move quickly along magnetic field lines. The intensities of sunward- and anti-sunward-moving particles in intervals a and b were similar (Fig. 2a), consistent with a corotating structure that traps particles between a

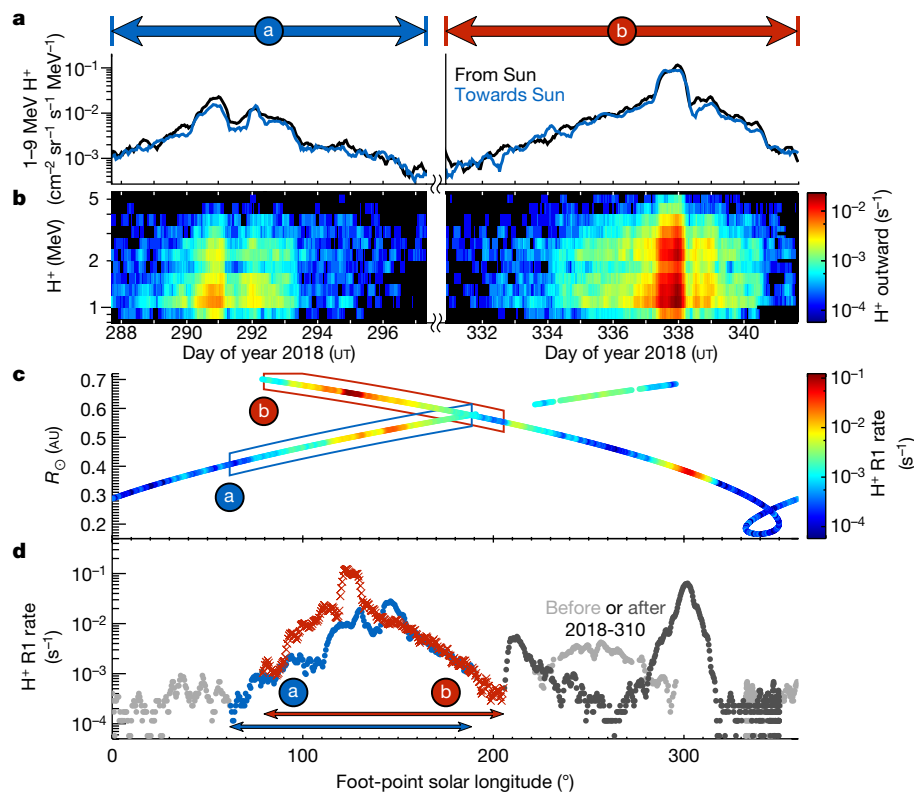


Fig. 2 | Recurring corotating energetic particle events. a–d, Corotating ion events observed in intervals a (blue) and b (red) represented by sunward/anti-sunward flux time series (a), a count-rate spectrogram (b), the flux as a

function of the magnetic foot point in Carrington longitude and radius from the Sun (c), and the 1–2 MeV H^+ rate versus foot-point longitude (d). In c, d, $v_{sw} = 350 \text{ km s}^{-1}$ is a nominal solar wind speed.

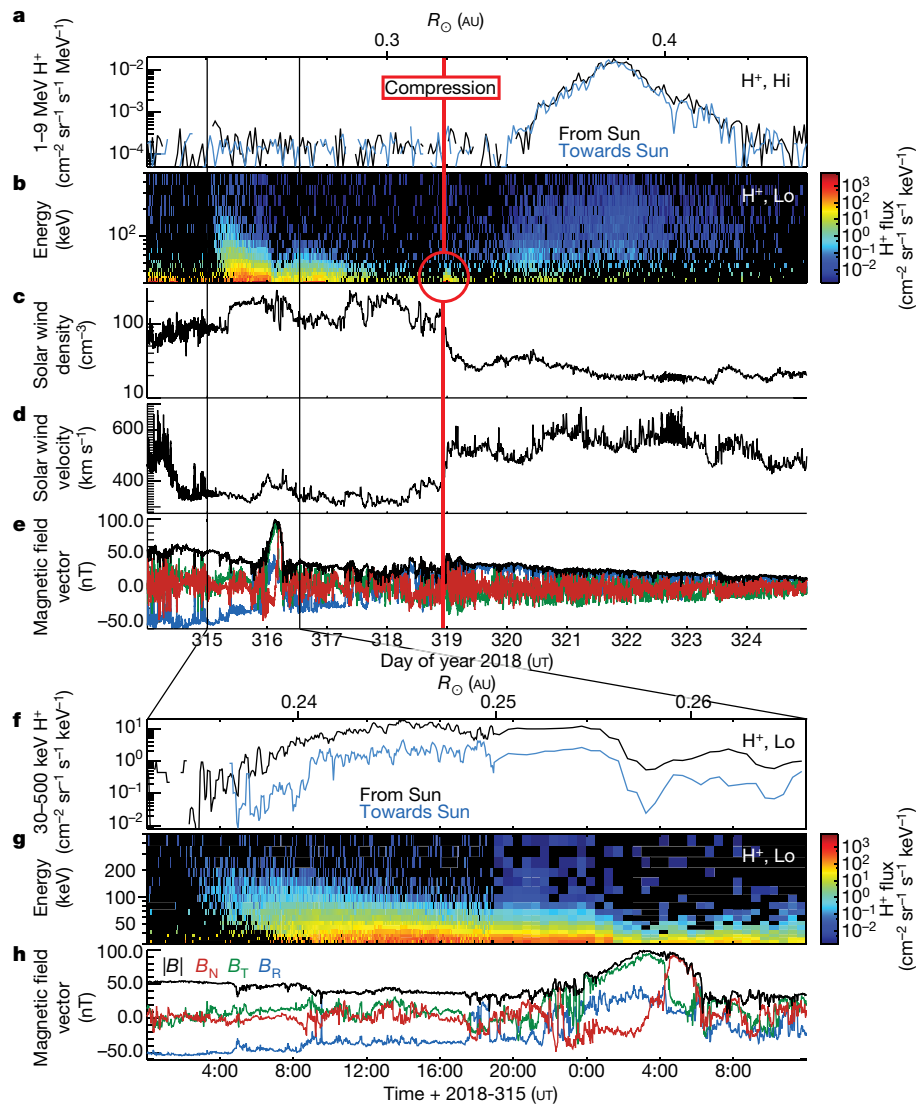


Fig. 3 | CME-related low-energy event and subsequent high-energy event.

a–e, Time series of primarily proton flux at >1 MeV (**a**), H^+ flux around 30–500 keV (**b**), solar wind density (**c**) and radial speed¹³ (**d**), and magnetic field

vector and magnitude¹⁴ (**e**) over interval **c**. **f–h**, Magnification of the dispersive SEP event and CME for H^+ flux around 30–500 keV.

source more distant than the spacecraft and the increasing magnetic field strength closer to the Sun. The particle acceleration probably occurs at reverse shocks, which typically form beyond about 2 AU from compressions in such CIRs.

The inbound leg towards perihelion 1 (which occurred at 0.17 AU, at 03:28 6 November 2018 UT) was extremely quiet from about 0.4 AU, providing an ideal opportunity for other PSP instruments^{13,14} to observe very quiet solar wind conditions with essentially no SEP-produced penetrating backgrounds. IS \odot IS began to observe lower-energy SEPs just before perihelion 1, which then increased. Figure 3 shows the events in interval **c**, including low-energy ions ahead of a coronal mass ejection (CME; Fig. 3b, f, g), the passage of a compression wave after it (Fig. 3c), and a subsequent higher-energy particle event (Fig. 3a).

IS \odot IS observations show an SEP event starting early on 2018-315 and extending to about when the CME arrived at PSP on 2018-316. Particle anisotropies (Fig. 3f) demonstrate that these particles are streaming outward from the Sun. The faster particles arrive first, characteristic of a ‘dispersive’ SEP event (Fig. 3g) with the differing arrival times giving an estimate of the distance along the magnetic field to the source of their acceleration. For the time–energy slope in Fig. 3g, we estimate a path length³ longer than that of the Parker spiral from PSP at about

0.25 AU, which might be explained by a longer path length associated with magnetic field ‘switchbacks’ observed by PSP in situ¹⁴.

Solar observations from the white-light coronagraph on the ‘A’ spacecraft of NASA’s Solar Terrestrial Relations Observatory (STEREO-A) indicate that the SEP-associated CME started lifting off from the Sun on 2018-314 at about 18:00 UT (Extended Data Fig. 1). Derivation of the CME speed from STEREO-A imaging (Extended Data Fig. 2) reveals that the CME was moving slowly (<400 km s⁻¹) from the Sun to PSP, very similar to the surrounding solar wind speed. By propagating this CME flux rope at a constant speed of 380 km s⁻¹ from near the Sun to PSP, we find good agreement with the in situ magnetic field observations. Preliminary analysis of this event using shock-modelling techniques¹⁵ suggests that there was probably no shock on the field lines well connected to PSP. However, a quasi-perpendicular sub-critical shock (Mach number <3) could have formed over an extended region of the flux rope and perhaps accelerated the protons measured by PSP (A. Kouloumvakos, private communication). This energetic particle event was not seen at any of the 1 AU spacecraft, so such small events may only be observable close to the Sun and therefore much more common than previously thought.

At the end of 2018-318, the solar wind speed increased from about 300 to about 500 km s⁻¹ (ref. 13), indicative of a strong dynamic pressure

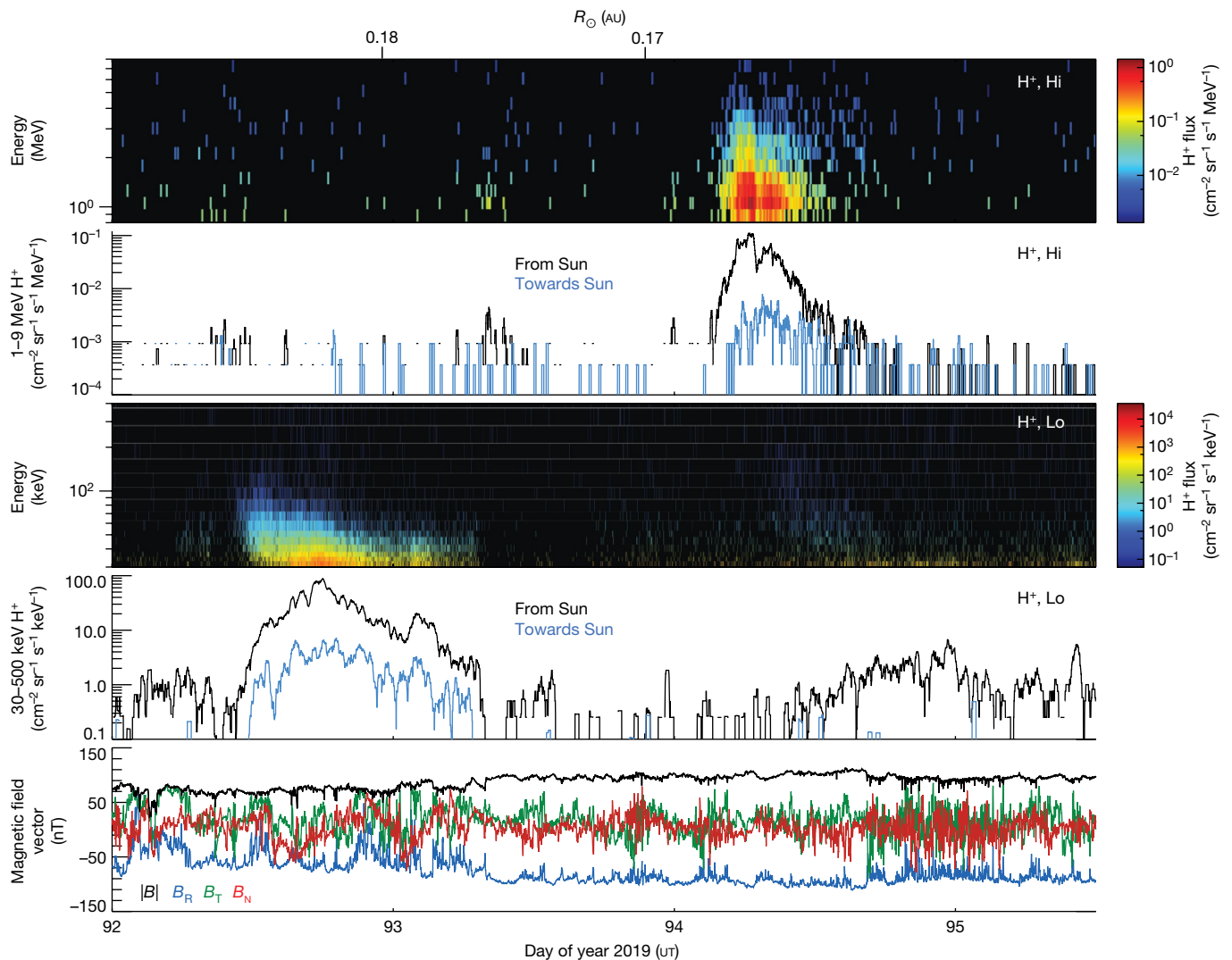


Fig. 4 | Pair of impulsive events near second perihelion. a–e. Two impulsive SEP events (during interval d) near the second perihelion of PSP ($<40R_{\odot}$) at higher energies (a, b), lower energies (c, d), and the magnetic field measured

during the events (e). $|B|$ is the magnitude of the magnetic field and B_R , B_T and B_N are the radial, transverse and normal components of the magnetic field, respectively.

wave in the solar wind. IS \odot IS observed a small enhancement in very-low-energy particles (<50 keV) as this compressional wave passed. This event is the first direct observation of local energization in the IS \odot IS observations. Shocks are not required for particle acceleration¹⁶ and plasma compressions can accelerate particles provided that the particles are able to propagate across, but remain close to, the compression¹⁷.

The large two-step increase in speed on 2018-318 shows that this compression wave was well on its way to steepening into a forward–reverse-shock pair, which probably accelerates the higher-energy (>1 MeV) particles observed from 2018-320 to 2018-324. This is not a CIR as in intervals a and b, because it has a much narrower range of foot-point longitudes (see H^+ count rate at about 300° in Fig. 2c, d) and does not recur, but instead indicates the interaction of a single fast solar wind stream, possibly associated with or even magnetically initiated by the preceding CME. In any case, as with CIR-associated particle events, the particle isotropy indicates that these ions are trapped on flux tubes, probably with a source more distant than PSP. In fact, at the time the second event was seen, about 1–6 days after the passage of the compression at PSP, the pressure front had expanded outward to heliocentric distances of about 0.6–2 AU, where the shocks probably formed.

Very near perihelion (about 35 solar radii, R_{\odot}) during PSP orbit 2 (interval d), IS \odot IS observed a unique pair of SEP events (Fig. 4). As PSP

is nearly corotational with the Sun near perihelion, the two events are magnetically connected to a common solar source $<5^\circ$ apart in longitude. First, on 2019-092 there was a low-energy dispersive event, probably associated with an impulsive source in the low corona. Two days later, on 2019-094, there was a different type of impulsive event, marked by a substantial increase in ions with >1 -MeV energy. Both events exhibit strong, persistent magnetic-field-aligned ions streaming away from the Sun.

The first event, starting on 2019-092, may be associated with disturbances also observed in extreme ultraviolet images from STEREO-A in the vicinity of active region AR2738, as well as multiple type-III radio bursts by both STEREO-A and FIELDS¹⁴, the magnetic field instrument aboard PSP. This small active region was about 70° off the nominal magnetic connection of PSP to the Sun. The fluxes of high-energy protons are near background, but we observed a substantial number of heavy high-energy ions and at low energies (about 30 keV per nucleon). He/H is about 20 times higher than the event on 2019-094, and increases in O and Fe abundances are even greater. These results suggest that this may be a ‘Z-rich’ event¹⁸; such events are relatively rare at 1 AU.

The second SEP event on 2019-094 also exhibits velocity dispersion and outward streaming, but has many fewer ions <1 MeV and a

substantial increase at >1 MeV. As with the event on 2019-092, there is potentially related radio and extreme ultraviolet activity in AR2738. However, the heavy ion abundances were similar to more typical SEP events. The magnetic field observed at PSP (Fig. 4e) between the two events was stronger and considerably smoother than before or after, indicating that this was probably a single, lower β (particle pressure/magnetic pressure) magnetic structure connecting the two events. Further, these observations indicate that processes inside 0.17 AU, as suggested by early multi-spacecraft studies in solar cycle 20, as well as later Helios and STEREO studies^{19–22}, enable fast, direct access of SEPs to a wide range of solar longitudes. Later studies that combined in situ data with solar source region observations showed that the smaller, longitudinally distributed SEP events are associated with multiple jet-like coronal emissions^{23,24} close to the source region as well as with more spatially extended eruptions²⁵.

IS \odot IS observed a rich array of energetic particle phenomena during PSP's first two orbits. Several of these events were not observed by 1 AU spacecraft, and so small events only observable close to the Sun may be much more common than previously thought. With these new data, we are well on the way to resolving the fundamental questions of the origin, acceleration, and transport of SEPs into the heliosphere. Over the next five years, as we head towards solar maximum, PSP will orbit progressively closer to the Sun, ultimately extending our exploration of these critical processes to less than 10 R_{\odot} .

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1811-1>.

1. Fox, N. J. et al. The Solar Probe Plus mission: humanity's first visit to our star. *Space Sci. Rev.* **204**, 7–48 (2016).
2. Mason, G. M. ^3He -rich solar energetic particle events. *Space Sci. Rev.* **130**, 231–242 (2007).
3. Desai, M. I. & Giacalone, J. Large gradual solar energetic particle events. *Living Rev. Sol. Phys.* **13**, 3 (2016).
4. Reames, D. V. *Solar Energetic Particles: A Modern Primer on Understanding Sources, Acceleration and Propagation* (Springer, 2017).

5. Cane, H. V., Richardson, I. G. & von Rosenvinge, T. T. A study of solar energetic particle events of 1997–2006: their composition and associations. *J. Geophys. Res. Space Phys.* **115**, A08101 (2010).
6. Wibberenz, G. & Cane, H. V. Multi-spacecraft observations of solar flare particles in the inner heliosphere. *Astrophys. J.* **650**, 1199–1207 (2006).
7. McComas, D. J. et al. Integrated Science Investigation of the Sun (ISIS): design of the energetic particle investigation. *Space Sci. Rev.* **204**, 187–256 (2016).
8. Kasper, J. C. et al. Solar Wind Electrons Alphas and Protons (SWEAP) investigation: design of the solar wind and coronal plasma instrument suite for Solar Probe Plus. *Space Sci. Rev.* **204**, 131–186 (2016).
9. Bale, S. D. et al. The FIELDS instrument suite for Solar Probe Plus. Measuring the coronal plasma and magnetic field, plasma waves and turbulence, and radio signatures of solar transients. *Space Sci. Rev.* **204**, 49–82 (2016).
10. Vourlidas, A. et al. The Wide-field Imager for Solar Probe Plus (WISPR). *Space Sci. Rev.* **204**, 83–130 (2016).
11. Pizzo, V. A three-dimensional model of corotating streams in the solar wind, 1. Theoretical foundations. *J. Geophys. Res.* **83**, 5563 (1978).
12. Gosling, J. Corotating and transient solar wind flows in three dimensions. *Annu. Rev. Astron. Astrophys.* **34**, 35–73 (1996).
13. Kasper, J. C. et al. Alfvénic velocity spikes and rotational flows in the near-Sun solar wind. *Nature* <https://doi.org/10.1038/s41586-019-1813-z> (2019).
14. Bale, S. D. et al. Highly structured slow solar wind emerging from an equatorial coronal hole. *Nature* <https://doi.org/10.1038/s41586-019-1818-7> (2019).
15. Kouloumvakos, A. et al. Connecting the properties of coronal shock waves with those of solar energetic particles. *Astrophys. J.* **876**, 80 (2019).
16. Chottoo, K. et al. The suprathermal seed population for corotating interaction region ions at 1 au deduced from composition and spectra of H^+ , He^{++} , and He^+ observed on Wind. *J. Geophys. Res. Space Phys.* **105**, 23107–23122 (2000).
17. Giacalone, J., Jokipii, J. R. & Kota, J. Particle acceleration in solar wind compression regions. *Astrophys. J.* **573**, 845–850 (2002).
18. Zwickl, R. D., Roelof, E. C., Gold, R. E., Krimigis, S. M. & Armstrong, T. P. Z-rich solar particle event characteristics 1972–1976. *Astrophys. J.* **225**, 281–303 (1978).
19. Reinhard, R. & Wibberenz, G. Propagation of flare protons in the solar atmosphere. *Sol. Phys.* **36**, 473–494 (1974).
20. Kallenrode, M. B. Particle propagation in the inner heliosphere. *J. Geophys. Res. Space Phys.* **98**, 19037–19047 (1993).
21. Richardson, I. G., von Rosenvinge, T. T. & Cane, H. V. The properties of solar energetic particle event-associated coronal mass ejections reported in different CME catalogs. *Sol. Phys.* **290**, 1741–1759 (2015).
22. Wiedenbeck, M. E. et al. Observations of solar energetic particles from ^3He -rich events over a wide range of heliographic longitude. *Astrophys. J.* **762**, 54 (2013).
23. Bučik, R. et al. ^3He -rich solar energetic particles from sunspot jets. *Astrophys. J. Lett.* **869**, 21 (2018).
24. Bučik, R., et al. Multi-spacecraft observations of recurrent ^3He -rich solar energetic particles. *Astrophys. J.* **786**, 71 (2014).
25. Nitta, N. V., Mason, G. M., Wang, L., Cohen, C. M. S. & Wiedenbeck, M. E. Solar sources of ^3He -rich solar energetic particle events in solar cycle 24. *Astrophys. J.* **806**, 235 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Data availability

All data used in this study is available to the public via NASA's Space Physics Data Facility (SPDF) at <https://spdf.gsfc.nasa.gov/>.

26. Howard, R. A. Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI). *Adv. Space Res.* **29**, 2017–2026 (2002).
27. Thernisien, A., Vourlidas, A. & Howard, R. Forward modeling of coronal mass ejections using STEREO/SECCHI data. *Sol. Phys.* **256**, 111 (2009).
28. Rouillard, A. P. et al. A solar storm observed from the Sun to Venus using the STEREO, Venus Express, and MESSENGER spacecraft. *J. Geophys. Res.* **114**, A07106 (2010).
29. Wood, B. & Howard, R. An empirical reconstruction of the 2008 April 26 coronal mass ejection. *Astrophys. J.* **702**, 901–910 (2009).
30. Isavnin, A. FRiED: a novel three-dimensional model of coronal mass ejections. *Astrophys. J.* **833**, 10 (2016).

Acknowledgements We are indebted to everyone who helped make the PSP mission possible. In particular, we thank all of the scientists, engineers, technicians, and administrative support people across all of the IS³IS institutions that produced and supported the IS³IS instrument suite and support its operations and the scientific analysis of its data. This work was supported as a part of the PSP mission under contract NNN06AA01C. S.D.B. acknowledges the support of the Leverhulme Trust Visiting Professorship programme and A.P.R. acknowledges financial support from the ANR project COROSHOCK ANR-17-CE31-0006-01 and from the ERC project SLOW_SOURCE – DLV-819189.

Author contributions D.J.M. is IS³IS Principal Investigator (PI) and led the data analysis and writing of the study. E.R.C. is IS³IS Deputy PI, helped develop EPI-Hi, and participated in the

data analysis. C.M.S.C. helped develop EPI-Hi and participated in the data analysis. A.C.C. helped develop EPI-Hi and participated in the data analysis. A.J.D. helped develop EPI-Hi and participated in the data analysis. M.I.D. participated in the data analysis. J.G. participated in the data analysis. M.E.H. helped develop EPI-Lo and participated in the data analysis. C.J.J. produced Figs. 3, 4 and participated in the data analysis. S.M.K. participated in the data analysis. A.W.L. helped develop EPI-Hi and participated in the data analysis. R.A.L. helped develop EPI-Hi and participated in the data analysis. O.M. participated in the data analysis. W.H.M. participated in the data analysis. R.L.M. led the development of EPI-Lo and participated in the data analysis. R.A.M. helped develop EPI-Hi and participated in the data analysis. D.G.M. helped develop EPI-Lo and participated in the data analysis. A.P. participated in the data analysis. J.S.R. helped develop EPI-Hi and participated in the data analysis. E.C.R. participated in the data analysis. N.A.S. led the development of the IS³IS Science Operations Center and participated in the data analysis. E.C.S. helped develop EPI-Hi and participated in the data analysis. J.R.S. led the development of the analysis tool, produced Figs. 1, 2, and participated in the data analysis. M.E.W. led the development of EPI-Hi and participated in the data analysis. S.D.B. is FIELDS PI and participated in the data analysis. J.C.K. is SWEAP PI and participated in the data analysis. A.W.C. helped develop SWEAP and participated in the data analysis. K.E.K. helped develop SWEAP and participated in the data analysis. R.J.M. helped develop FIELDS and participated in the data analysis. M.P. helped develop FIELDS and participated in the data analysis. M.L.S. helped develop SWEAP and participated in the data analysis. A.P.R. led the CME simulation work and participated in the data analysis.

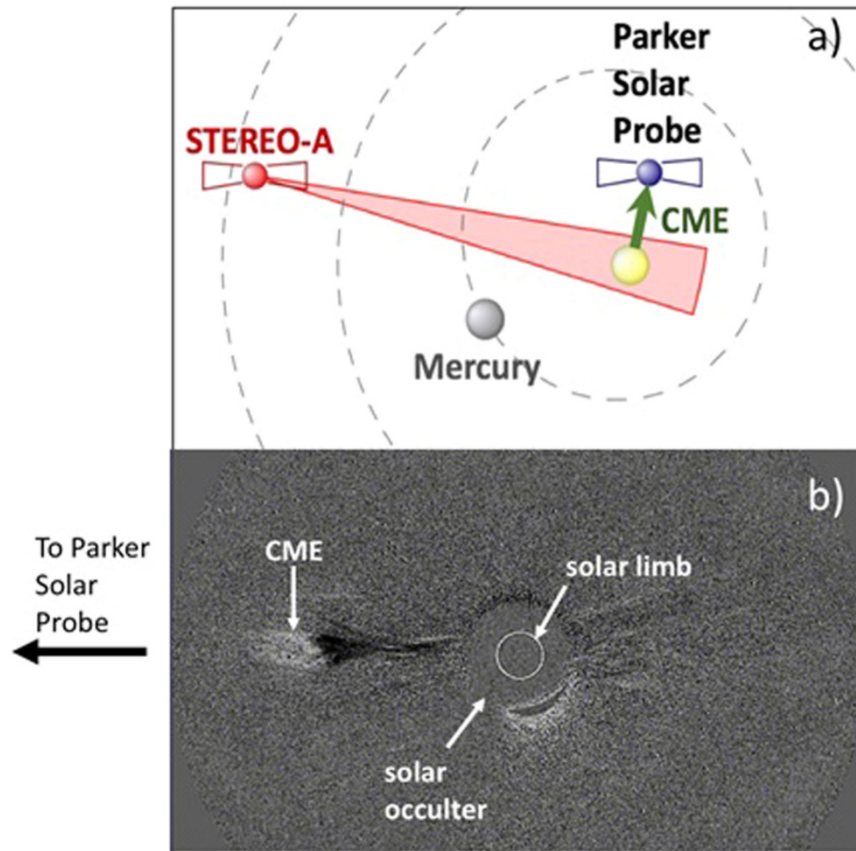
Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.J.M.

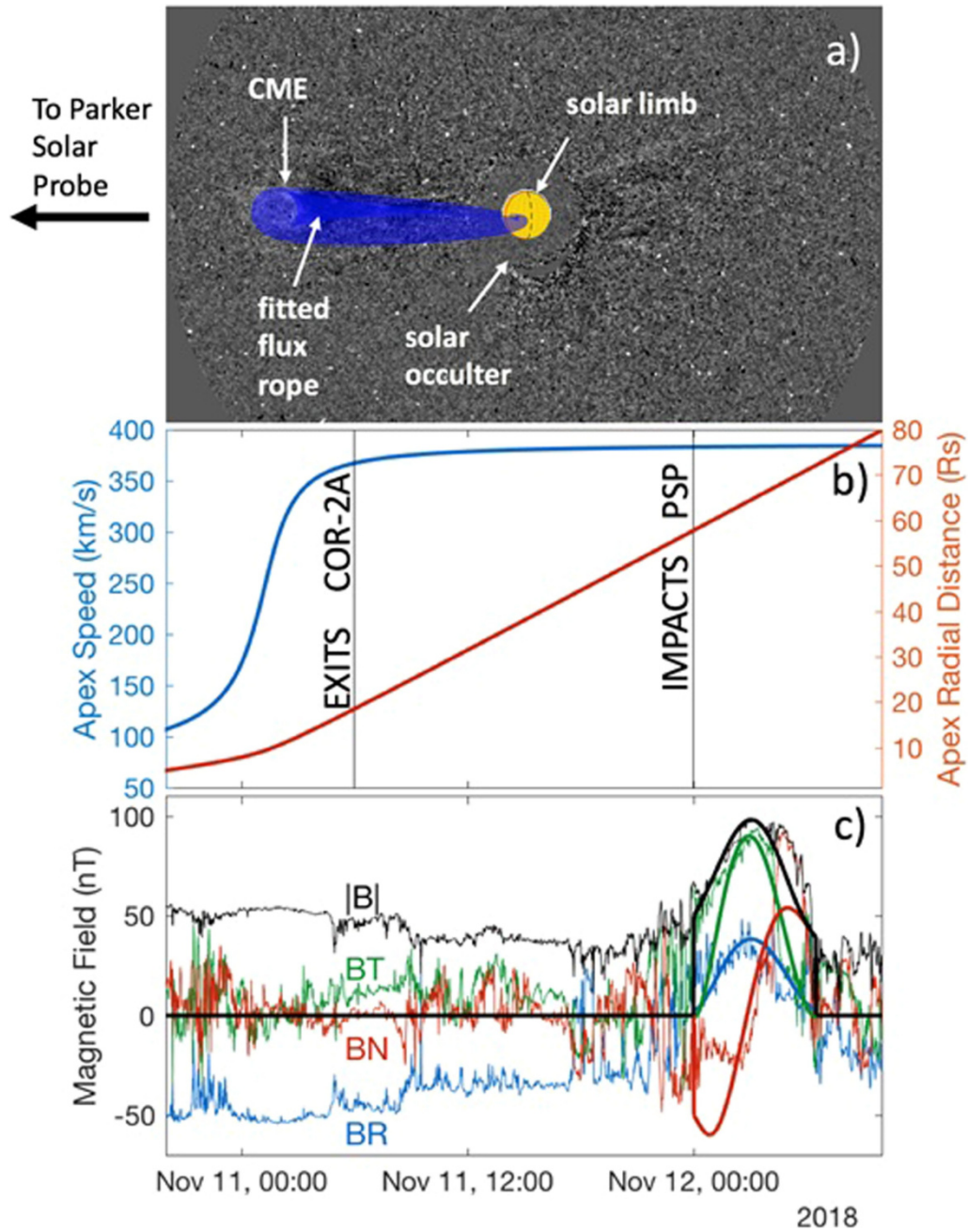
Peer review information *Nature* thanks Hazel Bain and Monica Laurenza for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Viewing geometry and observation of coronal mass ejection. **a**, A view of the ecliptic plane from solar north at 14:00 10 November 2018 UT showing the relative positions of STEREO-A and PSP. The dashed curves, from innermost to outermost, represent the orbits of Mercury, Venus and Earth. The red area shows the field of view of the COR-2 instrument onboard STEREO-A. A CME off the east limb of the Sun as viewed from STEREO-A would be roughly propagating towards PSP. This CME very gradually

entered the field of view of COR-2, part of the SECCHI suite of imaging instruments²⁶ aboard the STEREO spacecraft. **b**, A running-difference image of the CME taken at 02:39 UT on 11 November 2018 by COR-2A (a visible-light coronagraph), extending in the plane of the sky from $2R_{\odot}$ to $15R_{\odot}$, provided images during the entire acceleration phase of the CME. This CME entered COR-2A around 18:00 UT on 10 November 2018 and transited through the COR-2 field of view over about 12 h.



Extended Data Fig. 2 | CME model and comparison to magnetic field data.

a, The same as in Extended Data Fig. 1b but with superposed fitted flux-rope shape of the CME at 02:39 11 November 2018 UT when the CME had passed halfway through the COR-2A field of view. The CME is very weak and no shock-sheath structure can be identified in these images. The typical aspect of the CME in the image results from the line-of-sight integration of plasma distribution on a bent toroid such that its major axis is located in a plane containing the observing spacecraft (see very similar events in refs. ^{27,28}). **b**, The position (red) and speed (blue) of the apex of the flux-rope model was derived by iteratively comparing each synthetic image produced by the three-dimensional model with each available COR-2A image. A functional form (arctangent) was imposed for the flux rope's varying speed. The fitted CME structure assumed here is a bent toroid with an exponential increase of its cross-sectional area from foot point to apex as in ref. ²⁹. The speed was derived

by fitting a hyperbolic tangent to the modelled CME position. The speed increases rapidly from under 100 km s^{-1} at 18:00 10 November 2018 UT to over 350 km s^{-1} when it exited the COR-2A field of view at around 06:00 UT on 11 November. **c**, An internal magnetic field structure was expressed analytically inside the envelope of the fitted CME (smooth curves) as in ref. ³⁰, but here keeping a simple circular cross-section of the flux rope. By propagating this flux rope at a constant speed of 380 km s^{-1} from the time it exits the COR-2 field of view, we predict the CME reaches PSP on 12 November 2018. The predicted arrival time and the magnetic properties of the CME (thick smooth line) are in good agreement with those measured in situ by the FIELDs (magnetic field data; thin lines) and SWEAP instruments (not shown). We therefore conclude that the fitting procedure presented here provides a good description of the evolution of the CME from the upper corona to PSP.

Alfvénic velocity spikes and rotational flows in the near-Sun solar wind

<https://doi.org/10.1038/s41586-019-1813-z>

Received: 18 July 2019

Accepted: 17 October 2019

Published online: 4 December 2019

J. C. Kasper^{1,2*}, S. D. Bale^{3,4,5}, J. W. Belcher⁶, M. Berthomier⁷, A. W. Case², B. D. G. Chandran^{8,9}, D. W. Curtis⁴, D. Gallagher¹⁰, S. P. Gary¹¹, L. Golub², J. S. Halekas¹², G. C. Ho¹³, T. S. Horbury⁵, Q. Hu¹⁴, J. Huang¹, K. G. Klein^{15,16}, K. E. Korreck², D. E. Larson⁴, R. Liv⁴, B. Maruca^{17,18}, B. Lavraud¹⁹, P. Louarn¹⁹, M. Maksimovic²⁰, M. Martinovic¹⁵, D. McGinnis¹², N. V. Pogorelov¹⁴, J. D. Richardson⁶, R. M. Skoug¹¹, J. T. Steinberg¹¹, M. L. Stevens², A. Szabo²¹, M. Velli²², P. L. Whittlesey⁴, K. H. Wright²³, G. P. Zank¹⁴, R. J. MacDowall²¹, D. J. McComas²⁴, R. L. McNutt Jr¹³, M. Pulupa⁴, N. E. Raouafi¹³ & N. A. Schwadron^{8,9}

The prediction of a supersonic solar wind¹ was first confirmed by spacecraft near Earth^{2,3} and later by spacecraft at heliocentric distances as small as 62 solar radii⁴. These missions showed that plasma accelerates as it emerges from the corona, aided by unidentified processes that transport energy outwards from the Sun before depositing it in the wind. Alfvénic fluctuations are a promising candidate for such a process because they are seen in the corona and solar wind and contain considerable energy^{5–7}. Magnetic tension forces the corona to co-rotate with the Sun, but any residual rotation far from the Sun reported until now has been much smaller than the amplitude of waves and deflections from interacting wind streams⁸. Here we report observations of solar-wind plasma at heliocentric distances of about 35 solar radii^{9–11}, well within the distance at which stream interactions become important. We find that Alfvén waves organize into structured velocity spikes with duration of up to minutes, which are associated with propagating S-like bends in the magnetic-field lines. We detect an increasing rotational component to the flow velocity of the solar wind around the Sun, peaking at 35 to 50 kilometres per second—considerably above the amplitude of the waves. These flows exceed classical velocity predictions of a few kilometres per second, challenging models of circulation in the corona and calling into question our understanding of how stars lose angular momentum and spin down as they age^{12–14}.

The Parker Solar Probe (PSP) launched in August 2018 on a Delta IV Heavy rocket. The high energy of the launch, combined with a gravitational assist from Venus in September 2018, placed PSP into an eccentric orbit with a period of 147 days and a perihelion at a heliocentric distance of $r = 35.7R_{\odot}$ (R_{\odot} , solar radius) nearly a factor of two closer to the Sun than any previous mission⁴. This study uses observations made by instruments on the spacecraft during the first two encounters with the Sun, in November 2018 and April 2019. Whereas the instruments collect data at a low rate far from the Sun, the primary science collection at a high rate occurs during the encounter phase of each orbit at $r < 54R_{\odot}$ (0.25 AU). Encounter one (E1) lasted from 31 October to 12 November

2018, with the first perihelion occurring at 03:27 UT on 6 November. During these two encounters the longitude of PSP relative to the rotating surface of the Sun barely changed; PSP essentially dove down into, and then rose straight up from, a single narrow region above the Sun. E1 and E2 data thus describe a handful of specific solar-wind streams.

Nearly two million thermal-energy distribution functions of the solar-wind protons were recorded during E1, and more than three times that number during E2 (Fig. 1, Extended Data Fig. 1). From these distribution functions, the bulk properties of solar-wind protons—such as the velocity, density and temperature—are derived. Within any hour interval, the distribution of the radial solar-wind speed, V_{pr} , was strongly peaked at a

¹Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. ²Smithsonian Astrophysical Observatory, Cambridge, MA, USA. ³Physics Department, University of California, Berkeley, CA, USA. ⁴Space Sciences Laboratory, University of California, Berkeley, CA, USA. ⁵The Blackett Laboratory, Imperial College London, London, UK. ⁶Kavli Center for Astrophysics and Space Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁷Laboratoire de Physique des Plasmas, CNRS, Sorbonne Université, Ecole Polytechnique, Observatoire de Paris, Université Paris-Saclay, Paris, France. ⁸Department of Physics and Astronomy, University of New Hampshire, Durham, NH, USA. ⁹Space Science Center, University of New Hampshire, Durham, NH, USA. ¹⁰Heliophysics and Planetary Science Branch ST13, Marshall Space Flight Center, Huntsville, AL, USA. ¹¹Los Alamos National Laboratory, Los Alamos, NM, USA. ¹²Department of Physics and Astronomy, University of Iowa, IA, USA. ¹³Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA. ¹⁴Department of Space Science and Center for Space Plasma and Aeronomic Research, University of Alabama in Huntsville, Huntsville, AL, USA. ¹⁵Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA. ¹⁶Department of Planetary Sciences, University of Arizona, Tucson, AZ, USA. ¹⁷Department of Physics and Astronomy, University of Delaware, Newark, DE, USA. ¹⁸Bartol Research Institute, University of Delaware, Newark, DE, USA. ¹⁹Institut de Recherche en Astrophysique et Planétologie, CNRS, UPS, CNES, Université de Toulouse, Toulouse, France. ²⁰LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, Meudon, France. ²¹NASA/Goddard Space Flight Center, Greenbelt, MD, USA. ²²Department of Earth, Planetary and Space Sciences, University of California, Los Angeles, CA, USA. ²³Universities Space Research Association, Science and Technology Institute, Huntsville, AL, USA. ²⁴Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. *e-mail: jckasper@umich.edu

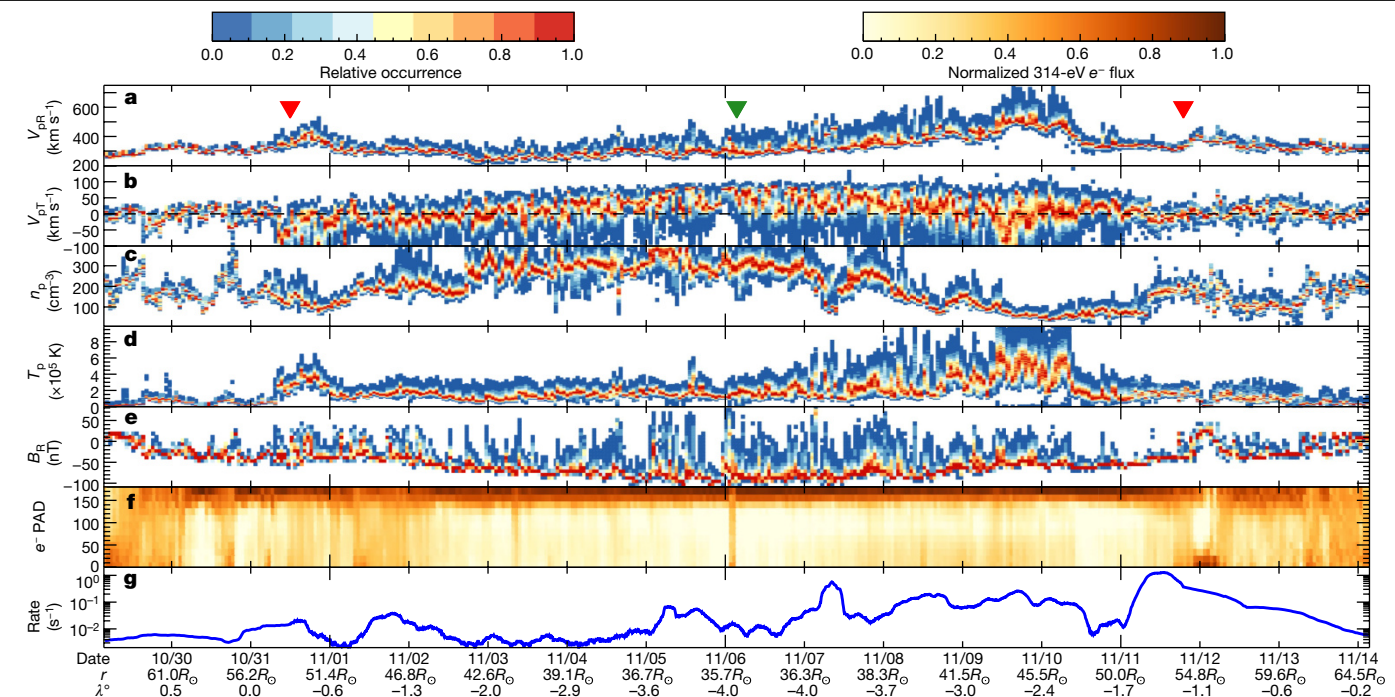


Fig. 1 | Overview of the first encounter of PSP with the Sun. **a**, Relative occurrence rate of the proton radial speed V_{PR} in one-hour intervals. Red triangles show the start and end of the high-rate data collection below $54R_{\odot}$ and the green triangle indicates a perihelion at $35.7R_{\odot}$. **b–f**, The same for V_{PT} in the solar equatorial plane (**b**), the proton number density n_p (**c**), the proton

temperature T_p (**d**), the radial component of magnetic field B_R (**e**), the electron pitch-angle distribution (PAD) (**f**) and the 20–200 keV proton rate (**g**). The date (month/day), distance r and latitude λ relative to the solar equator are indicated at daily intervals.

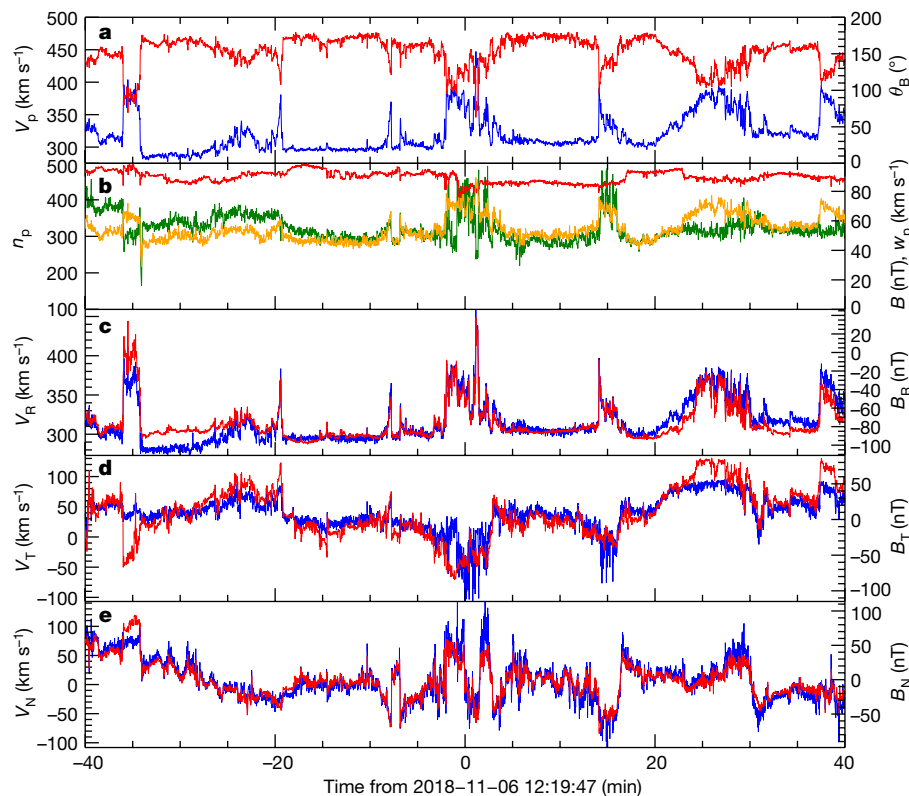


Fig. 2 | Solar-wind fluctuations near the closest approach. Near-Sun fluctuations meet the Alfvénic criteria, but are organized into structures and contain density enhancements. **a**, Magnitude of V_{PR} (blue) and angle θ_B of \mathbf{B} from the radial direction outwards. **b**, Magnitudes of n_p (green), B (red) and the proton thermal speed w_p (yellow). **c–e**, Variation of each vector component of the velocity (blue) and magnetic field (red) in the radial direction (R), the transverse direction in the solar equatorial plane (T) and the normal to R

and T (N). Since the orbit of PSP is within a few degrees of the solar equator, N points approximately north, perpendicular to the equatorial plane. There is a baseline solar-wind speed of about 300 km s^{-1} and jets where V_p jumps by about 100 km s^{-1} . The fluctuations are highly Alfvénic, with equal energy in the field and the flow, but they are organized into structures instead of being randomly distributed, and there is evidence of compressions.

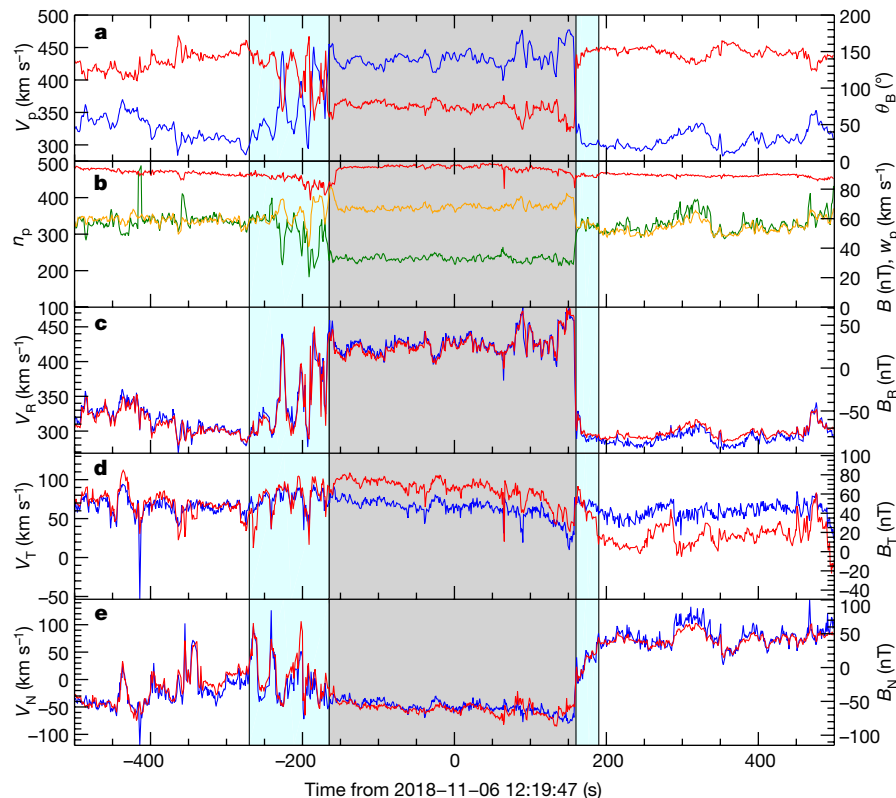


Fig. 3 | A closer look at a velocity spike. The formatting is as in Fig. 2, but focused on a single 1,000-s interval. The left blue region indicates the 105-s period in which PSP moved from the ambient plasma into the spike. The central core of the spike is indicated by the grey region and lasted for 325 s; it is

characterized by a steady but disturbed flow and a field with a large rotation in \mathbf{B} to $\theta_B \approx 70^\circ$ and a jump in flow to 343 km s^{-1} . The return from the core spike into ambient solar wind is marked by the second blue region and lasted 30 s.

minimum value, with a one-sided tail extending to larger V_{pR} values. V_{pR} reached its minimum of 200 km s^{-1} about a quarter of the way through E1 and then rose steadily to about 600 km s^{-1} . Numerical simulations and simple extrapolations of the observed photospheric magnetic field suggest that PSP spent all of E1 south of the global heliospheric current sheet, in a region with inward magnetic polarity ($B_R < 0$)¹⁵. Near the start and end of E1, PSP sampled slow wind from near the global heliospheric current sheet. Closer to the Sun, PSP first observed very slow wind and then fast wind, both of which are thought to emerge from a low-latitude coronal hole¹⁵. Below $40R_\odot$, V_{pT} (the transverse component of the proton velocity in the equatorial plane of the Sun) has a net positive value, which peaks at the closest approach. This flow may be the long-sought signature of plasma co-rotation in the corona. The density peaks in the slowest wind, at a value of approximately 400 cm^{-3} , about 50 times higher than values typically observed at 1 AU, as expected from mass conservation and spherical expansion. The proton temperature, T_p , and V_{pR} remain positively correlated¹⁶. At perihelion the protons are about 4 times hotter than protons with similar V_{pR} at 1 AU, consistent with radial scalings reported from earlier missions⁴. The radial component of the magnetic field, B_R , increases in magnitude with proximity to the Sun but unexpectedly changes sign many times. The pitch-angle (θ) distribution for electrons (that is, the number of electrons at a given energy as a function of their angle relative to \mathbf{B}) is a valuable diagnostic of these changes in the direction of \mathbf{B} . Here we show the pitch-angle distribution in a 22-eV-wide energy channel centred on 314 eV, well above the electron thermal energy. The sharp peak near 180° corresponds to the strahl, a beam of super-thermal electrons that travel away from the Sun along magnetic-field lines. Near the Sun, the strahl evolves towards small $\sin\theta$ values because of magnetic-moment conservation¹⁷. If the reversals in B_R seen by PSP result from the spacecraft's crossing between open field lines (connected to the Sun at only one end) with different

signs of B_R back at the Sun, then the strahl would flip between 180° and 0° each time B_R changed sign. Instead, every time B_R flips, the strahl maintains its 180° orientation, clearly indicating that the reversals in B_R are due to S-like bends in the magnetic-field lines (Extended Data Fig. 2). Closed field lines with both ends connected to the Sun and strahl travelling in both parallel and antiparallel directions to \mathbf{B} are seen during the arrival of a coronal mass ejection on 11 November at 23:50 UT, following an enhancement in the number of energetic particles¹⁸.

Figure 2 shows a timeseries of 80 min of observations several hours after perihelion, illustrating typical velocity and magnetic-field fluctuations. About half the time, \mathbf{B} points radially inwards towards the Sun and the velocity \mathbf{V} remains at a relatively constant 300 km s^{-1} . The remaining time includes seven distinct intervals in which \mathbf{B} rotates away from its radial-inward orientation, V_{pR} simultaneously jumps and \mathbf{V} also rotates, linking the one-sided tail in V_{pR} with the reversals in polarity seen in the E1 overview. These jumps in flow associated with rotations in \mathbf{B} and \mathbf{V} are similar to one-sided Alfvénic structures that were first seen farther from the Sun^{6,7}. The spikes seen by PSP are different in that they have larger amplitudes and are often associated with an increase in the proton density, n_p , indicating that the spikes have a non-Alfvénic component. The correlated variations in the components of \mathbf{B} and \mathbf{V} , their relative amplitudes and the constant value of $|\mathbf{B}|$ are consistent with large-amplitude, spherically polarized Alfvén waves propagating through the plasma in the anti-Sunward direction, similar to earlier observations^{5,19}. We can classify this wind stream (and indeed much of E1) as Alfvénic slow solar wind²⁰.

About 1,000 long-duration ($>10 \text{ s}$) and isolated velocity spikes with large rotations in \mathbf{B} were identified in E1 (about half as many were seen in E2.) Often the spikes can be separated chronologically into a core region with plasma conditions that are very different from those of the ambient solar wind but relatively constant, a comparatively short

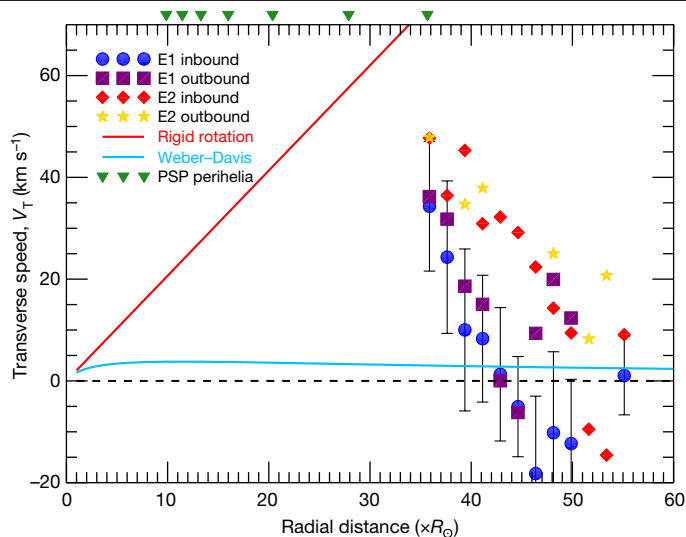


Fig. 4 | Large circulation of solar wind observed near the Sun. Averaged rotational (or azimuthal) flow, V_{PT} , over $1.75R_{\odot}$ intervals during E1 (inbound in blue, with error bars indicating the standard deviation from all observations; outbound in purple) and E2 (inbound in red; outbound in yellow) as a function of radial distance. Each symbol represents the average over at least 10,000 observations, and the values closest to perihelion are averaged over 60,000–230,000 observations. Error bars for E1 inbound show one standard deviation of the individual observations and are representative of the variation for the other three phases. The uncertainty in the mean of V_{PT} is much smaller than the marker size. Current and upcoming perihelia are shown with green triangles. Lines indicate no rotation (dashed), rigid co-rotation everywhere (red) and the axisymmetric Weber–Davis model (blue).

transition region on one side of the core, and a longer transition region on the other side containing large-amplitude fluctuations (see Fig. 3). During the 105-s transition at the beginning of this spike, the flow underwent seven large oscillations with an amplitude of 150 km s^{-1} , which possibly resulted from a Kelvin–Helmholtz instability.

Equally unexpected as the spikes and B_r reversals are the large-amplitude and sustained positive rotational velocities seen below $40R_{\odot}$ for E1 and $50R_{\odot}$ for E2 (Fig. 4). Net rotation has been reported farther from the Sun, but it was of the same order as instrument error and much smaller than the standard deviation in flow due to fluctuations and stream interactions^{8,21}. Here V_{PT} rises to 35 km s^{-1} (E1) and 50 km s^{-1} (E2). This is much greater than the variance from fluctuations including the velocity spikes, there is no evidence of stream interactions, and these values are much greater than the precision in the averaged flows (less than 0.5 km s^{-1}) and the absolute error in the flow due to a pointing error (less than 3 km s^{-1}) (see Methods). These are the first in situ observations of net rotational flow in the solar wind that are significantly above fluctuations and uncertainty.

Some level of rotational flow has always been expected in the solar wind near the Sun, as magnetic tension in the corona should force the plasma to rotate as the Sun spins. However, the large rotational velocities measured here greatly exceed the value calculated by the axisymmetric Weber–Davis model¹³, posing a major challenge to our understanding of the dynamics of the near-Sun solar wind. Determining the origin of these tangential flows will be essential for understanding how the Sun loses angular momentum and spins down as it ages^{12,14,22}. Further studies of the angular momentum should include magnetic

fields, waves and different ions. Future PSP orbits will clarify the extent to which these large rotational flows characterize other solar-wind streams. These orbits will also provide critical additional diagnostics of the state of the plasma, including turbulence, velocity spikes, temperature anisotropy and particle velocity distribution functions, at heliocentric distances as small as $9.86R_{\odot}$.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1813-z>.

- Parker, E. N. Dynamics of the interplanetary gas and magnetic fields. *Astrophys. J.* **128**, 664–676 (1958).
- Gringauz, K. I., Bezrokh, V. V., Ozerov, V. D. & Rybchinskii, R. E. A study of the interplanetary ionized gas, high-energy electrons and corpuscular radiation from the Sun by means of the three-electrode trap for charged particles on the second Soviet cosmic rocket. *Sov. Phys. Dokl.* **5**, 361–364 (1960).
- Bonetti, A., Bridge, H. S., Lazarus, A. J., Rossi, B. & Scherb, F. Explorer 10 plasma measurements. *J. Geophys. Res.* **68**, 4017–4063 (1963).
- Marsch, E. et al. Solar wind protons – three-dimensional velocity distributions and derived plasma parameters measured between 0.3 and 1 AU. *J. Geophys. Res.* **87**, 52–72 (1982).
- Belcher, J. W. & Davis, L. Jr Large-amplitude Alfvén waves in the interplanetary medium, 2. *J. Geophys. Res.* **76**, 3534–3563 (1971).
- Gosling, J. T., McComas, D. J., Roberts, D. A. & Skoug, R. M. A one-sided aspect of Alfvénic fluctuations in the solar wind. *Astrophys. J. Lett.* **695**, 213–216 (2009).
- Horbury, T. S., Matteini, L. & Stansby, D. Short, large-amplitude speed enhancements in the near-Sunfast solar wind. *Mon. Not. R. Astron. Soc.* **478**, 1980–1986 (2018).
- Pizzo, V. et al. Determination of the solar wind angular momentum flux from the HELIOS data – an observational test of the Weber and Davis theory. *Astrophys. J.* **271**, 335–354 (1983).
- Fox, N. J. et al. The solar probe plus mission: humanity's first visit to our star. *Space Sci. Rev.* **204**, 7–48 (2016).
- Kasper, J. C. et al. Solar Wind Electrons Alphas and Protons (SWEAP) investigation: design of the solar wind and coronal plasma instrument suite for Solar Probe Plus. *Space Sci. Rev.* **204**, 131–186 (2016).
- Bale, S. D. et al. The FIELDS instrument suite for Solar Probe Plus. Measuring the coronal plasma and magnetic field, plasma waves and turbulence, and radio signatures of solar transients. *Space Sci. Rev.* **204**, 49–82 (2016).
- Schatzman, E. A theory of the role of magnetic activity during star formation. *Ann. d'Astr.* **25**, 18–29 (1962).
- Weber, E. J. & Davis, L. Jr The angular momentum of the solar wind. *Astrophys. J.* **148**, 217–227 (1967).
- Finley, A. J., Matt, S. P. & See, V. The effect of magnetic variability on stellar angular momentum loss. I. The solar wind torque during sunspot cycles 23 and 24. *Astrophys. J.* **864**, 125 (2018).
- Bale, S. D. et al. Highly structured slow solar wind emerging from an equatorial coronal hole. *Nature* <https://doi.org/10.1038/s41586-019-1818-7> (2019).
- Elliott, H. A., Henney, C. J., McComas, D. J., Smith, C. W. & Vasquez, B. J. Temporal and radial variation of the solar wind temperature–speed relationship. *J. Geophys. Res. Space Phys.* **117**, A09102 (2012).
- Pilipp, W. G. et al. Characteristics of electron velocity distribution functions in the solar wind derived from the helios plasma experiment. *J. Geophys. Res. Space Phys.* **92**, 1075–1092 (1987).
- McComas, D. M. et al. Probing the energetic particle environment near the Sun. *Nature* <https://doi.org/10.1038/s41586-019-1811-1> (2019).
- Vasquez, B. J. & Hollweg, J. V. Formation of arc-shaped Alfvén waves and rotational discontinuities from oblique linearly polarized wave trains. *J. Geophys. Res.* **101**, 13527–13540 (1996).
- Bruno, R. & Carbone, V. The solar wind as a turbulence laboratory. *Living Rev. Sol. Phys.* **10**, 2 (2013).
- Richardson, I. G. Solar wind stream interaction regions throughout the heliosphere. *Living Rev. Sol. Phys.* **15**, 1 (2018).
- Axford, W. I. The solar wind. *Sol. Phys.* **100**, 575–586 (1985).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Data collection and analysis

The data presented here were collected over the course of the first two encounters of the Sun by PSP in November 2018 and April 2019. This study makes use of all of the in situ instruments on the spacecraft. The thermal plasma properties are measured by the PSP SWEAP (Solar Wind Electrons Alphas and Protons) instrument suite¹⁰, including the Solar Probe Cup (SPC) and the Solar Probe Analyzers (SPAN) for electron and ion plasma data. Magnetic-field data from the outboard FIELDS magnetometer were also used^{11,15}, along with energetic-particle rates as seen by ISOIS¹⁸. SPC measures the reduced distribution function of ionized hydrogen and helium and the two-dimensional flow angles of the ions as a function of energy/charge. These measurements were performed at least once per second and typically more than four times per second throughout the encounter phase of each orbit (below 0.25 AU or 54 R_{\odot}). This paper uses moments of the entire SPC proton distribution function to calculate the total effective proton velocity, density and radial component of the temperature. While the SPAN ion sensor generally did not view the peak of the proton velocity distribution, the overlapping regions seen by SPAN and SPC were compared to confirm that there were no gross offsets in the calibration or the derived plasma properties such as the velocity; this technique will be more accurate when the solar wind flows into SPAN closer to the Sun. Observations of electrons with a central energy of 314 eV and a width of 22 eV by the two SPAN electron sensors were combined, along with the FIELDS determination of the magnetic-field direction, to create the electron pitch-angle distributions.

All data are being archived and will be available for download at the NASA Space Physics Data Facility in November 2019 (<https://spdf.gsfc.nasa.gov/>). Additional SWEAP data and information are available at the SWEAP web page (<http://sweap.cfa.harvard.edu/>). Data were analysed and graphics were developed in IDL (Interactive Data Language).

Statistics

The distributions of plasma properties in Fig. 1 and Extended Data Fig. 1 were produced with a time resolution of 1 h. During the encounters, the time resolution of the plasma instrument ranged from slightly more than one measurement per second to more than four measurements per second, so each column represents the distribution of approximately 3,600–14,400 measurements. All error bars indicate one standard deviation of the measurements from the mean. At least 10,000—and generally more than 80,000—observations were used to calculate the mean transverse flow V_{\perp} in Fig. 4.

Estimates of uncertainty

Here we discuss the absolute accuracy of SPC ion measurements. As verified by ground testing, the absolute accuracy for V_{PR} is less than 0.01% over a measurable range of approximately 119 km s⁻¹ to 1,065 km s⁻¹. The absolute accuracy in temperature is similarly small over a measurable range of approximately 7.3 kK to 21.1 MK (that is, proton thermal speeds of 11 km s⁻¹ to 600 km s⁻¹). Speeds and temperatures at the extremes of these ranges are subject to systematic considerations, but no such measurements are presented here. The accuracy of the density measurement is determined by comparison with the plasma frequency as observed by FIELDS¹¹. Thus, the absolute accuracy of the SPC density measurement is estimated to be about 1% and is no worse than 3%. The absolute accuracy for off-radial flow components is verified via spacecraft roll manoeuvres about the SPC symmetry axis. For solar-wind fluxes typical of the first two encounters, the uncertainty associated with this calibration corresponds to a typical absolute accuracy of about 0.5°. For a solar wind of 400 km s⁻¹ this corresponds to an expected error in V_{PT} of 3–4 km s⁻¹, which is much smaller than the net rotational flow observed.

Signatures of Alfvénic fluctuations

In discussing Fig. 2, we stated that the correlation of fluctuations in the components of \mathbf{B} and \mathbf{V} were generally indicative of outward-propagating Alfvén waves. We consider the vector waves or fluctuations $\Delta\mathbf{V}$ and $\Delta\mathbf{B}$ superimposed on a steady background of \mathbf{B}_0 and \mathbf{V}_0 , respectively. In the long-wavelength fluid magnetohydrodynamic limit, Alfvén waves propagate exactly in parallel or antiparallel directions to \mathbf{B}_0 , are dispersionless and do not compress the plasma, and there is a simple linear relationship of $\Delta\mathbf{B} = \pm D_A \Delta\mathbf{V}$, where $D_A = (n_p + 4n_\alpha)^{0.5} \theta / 21.8$ (in units of nT km⁻¹ s; densities are in units of cm⁻³) and $\theta = (1 - \beta_{\parallel} + \beta_{\perp})^{-0.55}$. Here θ is a correction for thermal pressure anisotropy, where β_{\parallel} is the ratio of the parallel plasma pressure to the magnetic pressure and β_{\perp} is the ratio of the perpendicular plasma pressure to the magnetic pressure. For this period we find on average $n_p = 220 \text{ cm}^{-3}$, $\beta_{\parallel} = 0.202$ and $\beta_{\perp} = 0.315$. SPC and SPAN were not configured optimally to measure the ionized helium abundance n_α , so assuming the typical range $0.5\% < n_\alpha/n_p < 4.5\%$, we expect $D_A = 0.68\text{--}0.74 \text{ nT km}^{-1} \text{ s}$. We find D_A for the R, T, N components to be 0.71, 1.09 and 0.70 nT km⁻¹ s, respectively, so the R and N components are exactly within the expected range and the fluctuations in the T direction are about 33% higher (it is typical for D_A to be different for each component of the velocity⁵). We then used the calculated value of D_A to rescale the range of the vector components of \mathbf{B} , so they should overlap with \mathbf{V} if the fluctuations were purely Alfvénic. The sign of the relation between $\Delta\mathbf{B}$ and $\Delta\mathbf{V}$ is given by the sign of $-\mathbf{k} \cdot \mathbf{B}_0$, where \mathbf{k} is the wavevector and gives the direction of propagation, and \mathbf{B} is an average direction of the field over a long time scale. The ambient direction of the magnetic field outside the large-amplitude fluctuations points towards the Sun and the correlations are overwhelmingly positive, meaning that we are seeing outward-propagating waves.

Identification of velocity spikes

Isolated velocity spikes were identified by looking for all intervals in each encounter in which the orientation of the magnetic field started in the quiet configuration pointed towards the Sun within 30°, rotated more than 45° away from the quiet configuration for at least 10 s, and then returned to the original direction. Candidate events were then examined manually to identify starting and ending times.

Acknowledgements The SWEAP Investigation and this study are supported by the PSP mission under NASA contract NNN06AA01C. The SWEAP team expresses its gratitude to the scientists, engineers and administrators who have made this project a success, both within the SWEAP institutions and from NASA and the project team at JHU/APL. J.C.K. acknowledges support from the 2019 Summer School at the Center for Computational Astrophysics, Flatiron Institute. The Flatiron Institute is supported by the Simons Foundation. S.D.B. acknowledges the support of the Leverhulme Trust Visiting Professorship programme. T.S.H. was supported by UK STFC ST/S000364/1.

Author contributions J.C.K. is the SWEAP Principal Investigator (PI) and led the data analysis and writing of this Article. S.D.B. is the FIELDS PI and a SWEAP Co-Investigator and provided the magnetic-field observations. J.W.B. leads the US group that developed the solar-wind Faraday cup, and provided guidance on identifying Alfvénic fluctuations. M.B. provided a pre-amplifier ASIC used within the SPAN electron instruments. A.W.C. is the SPC instrument scientist and ensured that the instrument met its performance requirements and was calibrated. B.D.G.C. contributed to the theoretical calculations and the writing of the manuscript. D.W.C. managed the effort at UCB. D.G. was the institutional lead at NASA MSFC, responsible for materials testing and calibration of SPC. S.P.G. provided recommendations on measurement requirements to detect instabilities. L.G. provided related solar observations and results. J.S.H. contributed to the analysis of the electron observations and to the writing of the manuscript. G.C.H. provided a time-of-flight ASIC to reduce the size and power of the SPAN ion instrument. T.S.H. participated in the analysis of the Alfvénic spikes. Q.H. identified magnetic flux ropes. K.G.K. contributed to the writing of the manuscript and provided warm-plasma growth rate calculations. K.E.K. led the SWEAP Science Operations Center and coordinated observing plans between the instruments and the project. M.V. contributed to the writing of the manuscript and the discussion on the relationship between Alfvénic fluctuations and angular momentum. D.E.L. is the institutional lead at Berkeley, responsible for the implementation of the SPAN instruments and the SWEAP Electronics Module suite-wide computer. R.L. is the SPAN ion instrument scientist. B.L. identified flux ropes and other signatures of coronal mass ejections in the data. P.L. coordinated solar furnace testing of the SPC materials before launch. M. Maksimovic performed the absolute calibration of the density measurements. M. Martinovic evaluated the quality of the velocity distribution functions. N.V.P. carried out numerical

simulations. J.D.R. contributed to the design of the Faraday cup. R.M.S. helped to interpret the electron pitch-angle distributions. J.T.S. identified potential field rotation causes. M.L.S. provided the overall data pipeline for SWEAP and SPC high-level data products. A.S. estimated the location of the heliospheric current sheet. P.L.W. set up the SPC calibration at MSFC and then became SPAN electron instrument scientist at Berkeley. K.H.W. arranged the SPC calibration at MSFC. G.P.Z. leads the SWEAP theory team. R.J.M. leads the FIELDS fluxgate magnetometer. D.J.M. is the IS \odot S PI and provided the energetic-particle data. R.L.M. leads the EPI-Lo energetic-particle instrument. M.P. is the FIELDS SOC lead. N.E.R. is the PSP Project Scientist and reviewed jets and similar coronal transients. N.A.S. runs the

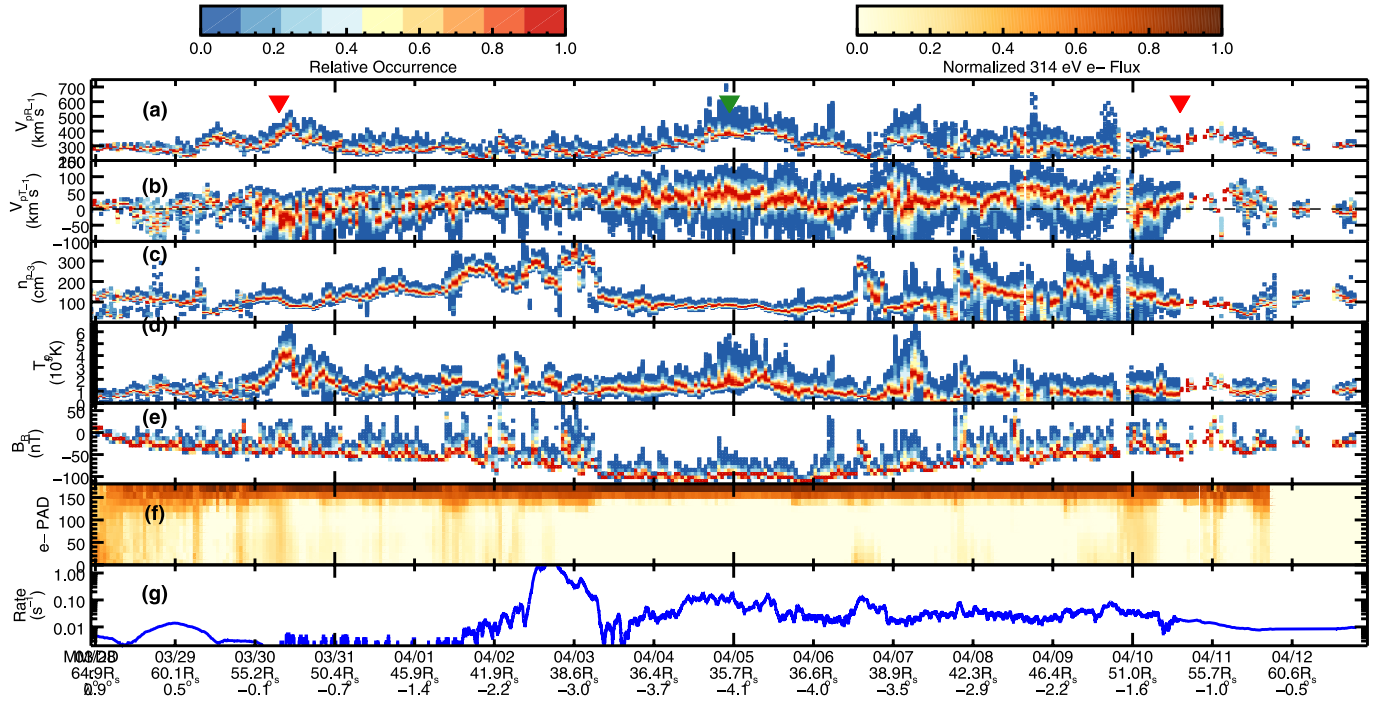
IS \odot S Science Operations Center. All authors participated in planning the observations and data collection, reviewed and discussed the observations, and read, provided feedback and approved the contents of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.C.K.

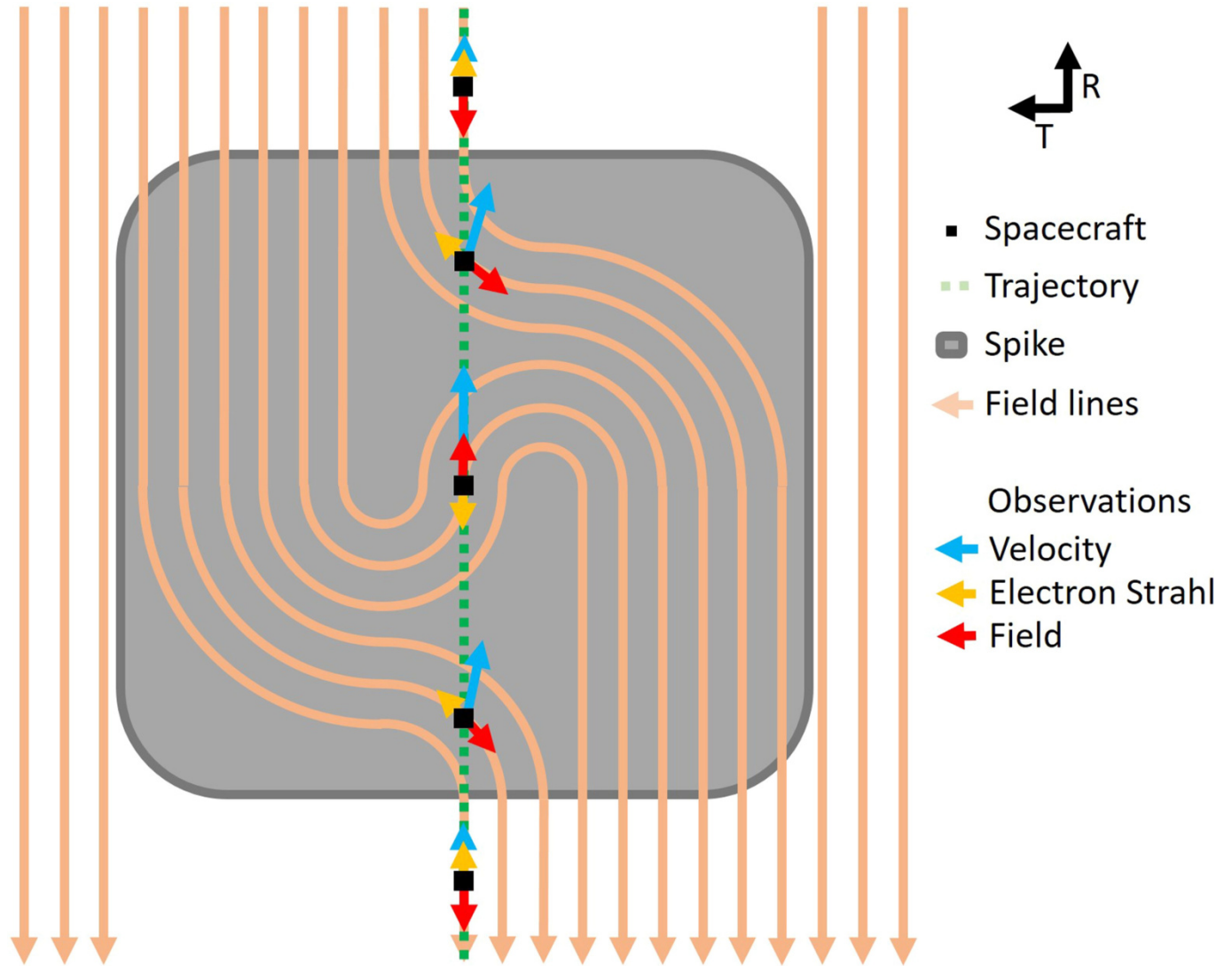
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Overview of the second PSP encounter with the Sun.

The figure is in the same format as Fig. 1. Spikes in the velocity are again seen to be coincident with the magnetic-field reversals, but the jump in the speed is

smaller, probably because the Alfvén speed was lower in E2 than E1. The density at perihelion is substantially lower.



Extended Data Fig. 2 | Schematic of an S-shaped magnetic structure creating a field reversal, heat-flux reversal and a spike in velocity. This figure illustrates the possible geometry of an S-shaped propagating Alfvénic disturbance (grey box) and how it would appear to the spacecraft (black square) as it flew through the spike on the green trajectory. The pink lines with arrows indicate the configuration of the magnetic field, with all field lines ultimately pointing back to the Sun. Arrows at each black square indicate the vector velocity (blue), electron strahl (orange) and magnetic field (red) seen by the spacecraft. If this was a purely Alfvénic structure, then the spike would move away from the Sun in an antiparallel direction to \mathbf{B} at the local Alfvén

speed, C_A . In the frame of the spike, the shape of the structure would be static, with plasma flowing in along field lines on the upper left and through the spike and emerging at the lower right, always flowing at C_A . In the frame of the spacecraft, the constant flow along field lines in the propagating spike frame would translate into a radial increase of V by C_A when \mathbf{B} is perpendicular to the R direction, and a maximum jump of $2C_A$ when \mathbf{B} is completely inverted. Because the heat flux flows away from the Sun along magnetic field lines, it would rotate so as to always be antiparallel to \mathbf{B} and appear locally to be flowing back to the Sun at the centre of this disturbance.

Near-Sun observations of an F-corona decrease and K-corona fine structure

<https://doi.org/10.1038/s41586-019-1807-x>

Received: 28 June 2019

Accepted: 8 November 2019

Published online: 4 December 2019

R. A. Howard^{1*}, A. Vourlidas², V. Bothmer³, R. C. Colaninno¹, C. E. DeForest⁴, B. Gallagher¹, J. R. Hall⁵, P. Hess¹, A. K. Higginson², C. M. Korendyke¹, A. Kouloumvakos⁶, P. L. Lamy⁷, P. C. Liewer⁸, J. Linker⁹, M. Linton¹, P. Penteado⁵, S. P. Plunkett⁹, N. Poirier⁶, N. E. Raouafi², N. Rich¹, P. Rochus¹⁰, A. P. Rouillard⁶, D. G. Socker¹, G. Stenborg¹, A. F. Thernisien¹ & N. M. Viall¹¹

Remote observations of the solar photospheric light scattered by electrons (the K-corona) and dust (the F-corona or zodiacal light) have been made from the ground during eclipses¹ and from space at distances as small as 0.3 astronomical units^{2–5} to the Sun. Previous observations^{6–8} of dust scattering have not confirmed the existence of the theoretically predicted dust-free zone near the Sun^{9–11}. The transient nature of the corona has been well characterized for large events, but questions still remain (for example, about the initiation of the corona¹² and the production of solar energetic particles¹³) and for small events even its structure is uncertain¹⁴. Here we report imaging of the solar corona¹⁵ during the first two perihelion passes (0.16–0.25 astronomical units) of the Parker Solar Probe spacecraft¹³, each lasting ten days. The view from these distances is qualitatively similar to the historical views from ground and space, but there are some notable differences. At short elongations, we observe a decrease in the intensity of the F-coronal intensity, which is suggestive of the long-sought dust free zone^{9–11}. We also resolve the fine-scale plasma structure of very small eruptions, which are frequently ejected from the Sun. These take two forms: the frequently observed magnetic flux ropes^{12,16} and the predicted, but not yet observed, magnetic islands^{17,18} arising from the tearing-mode instability in the current sheet. Our observations of the coronal streamer evolution confirm the large-scale topology of the solar corona, but also reveal that, as recently predicted¹⁹, streamers are composed of yet smaller substreamers channelling continual density fluctuations at all visible scales.

The Parker Solar Probe (PSP) carries an imaging instrument, the Wide-field Imager for Solar Probe (WISPR)¹⁵. The inset in Fig. 1a shows a WISPR inner telescope (WISPR-I) image taken on 6 November 2018 at the first perihelion. The Sun is 13.5° to the left of the image and the width is about 40°. The locus of points at the apex of the contours defines the photometric axis of the F-corona. While most observations of the F-corona (or zodiacal light) have been taken from 1 astronomical unit (AU) away from the Sun, two spacecraft, Helios A and B, each carrying the Zodiacal Light Experiment⁵, orbited the Sun from 0.3 to 1.0 AU, one observing above the ecliptic plane and the other below. They measured the intensity I of the zodiacal light from varying heliocentric distances and found²⁰ that it increases towards the Sun according to $I \propto R_0^{-n}$, where $n = 2.3 \pm 0.1$ and R_0 is the radius of the Sun. The upper and lower limits were recorded at small and large elongations from the Sun, respectively, and were independent of the ecliptic longitude of the observer. The Sun Earth Connection Coronal and Heliospheric Investigation⁴ (SECCHI) heliospheric imagers HI-1²¹, onboard the STEREO spacecraft²²

orbiting the Sun at approximately 1 AU, observed the corona at elongations ranging from 0.07 to 0.45 AU (5°–24°) from the Sun. An analysis of intensities²³ of the photometric axis of the F-corona from 2007 to 2014 found the exponent, for the entire elongation range covered by the HI-1 instrument, to be 2.31. Moreover, the analysis performed on restricted elongation ranges²³ showed an identical tendency to the Helios results for the intensity gradient to increase towards the Sun ($n = 2.29 \pm 0.10$).

Figure 1a displays a log–log plot of a sample of F-coronal intensity profiles in units of mean solar brightness (MSB) along its photometric axis as measured by WISPR-I between 15° and 50° elongation from the centre of the Sun. The sample comprises data from five different heliocentric distances of the PSP spacecraft (0.336 AU to 0.166 AU) obtained during the orbit inbound to the first perihelion. Colour is used to distinguish the plots. For clarity we plot only these five positions, but all the profiles during the encounter are similar. These five profiles are normalized to the maximum intensity at 30° elongation

¹US Naval Research Laboratory, Washington, DC, USA. ²Johns Hopkins University, Applied Physics Laboratory, Laurel, MD, USA. ³Institut für Astrophysik, University of Göttingen, Göttingen, Germany. ⁴Southwest Research Institute, Boulder, CO, USA. ⁵Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. ⁶L'Institut de Recherche en Astrophysique et Planétologie, Toulouse, France. ⁷Laboratoire Atmosphères, Milieux et Observations Spatiales, CNRS and UVSQ, Guyancourt, France. ⁸Predictive Science Inc., San Diego, CA, USA. ⁹National Aeronautics and Space Administration, Washington, DC, USA. ¹⁰University of Liège, Liège, Belgium. ¹¹Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD, USA. *e-mail: Russ.Howard@nrl.navy.mil

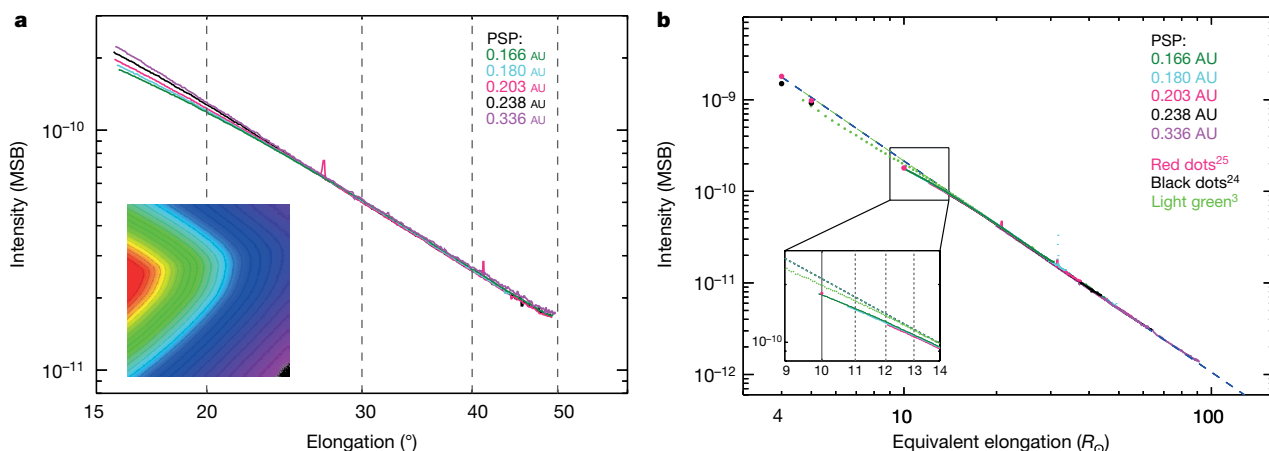


Fig. 1 | Intensity plots along the photometric axis of the F-corona.

a, Observed intensities from WISPR-I for five heliocentric distances as a function of elongation (degrees) scaled to the same value at 30° elongation.

The inset in **a** shows an image of the F-corona taken on 6 November 2018. **b**, Observed intensities from both telescopes for five heliocentric distances as a function of elongation (solar radii). See text for further explanation.

to reveal the behaviour of the profiles for the various PSP heliocentric distances. Clearly, at larger elongations, the curves have exactly the same slope, and at shorter elongations ($<20^\circ$), the intensity decreases with decreasing PSP distance, with the top plot (pink) for when PSP is the furthest from the Sun and the bottom plot (dark green) the closest. Figure 1b shows the intensity profiles at the same five PSP distances for both telescopes, but plotted against elongations converted to R_\odot . The conversion to R_\odot was performed by dividing the elongation by half the angular size of the Sun at the respective PSP distance. We note that the curves all overlie each other now, even the decreases seen in Fig. 1a. The small upward ticks are due to bright stars. The dashed blue line in Fig. 1b shows the linear fit to the F-coronal intensities for elongations between $20R_\odot$ and $77R_\odot$ ($n = 2.31$), a result identical to that obtained from both earlier observations^{20,23}. For comparison, historical data^{3,24,25} have been added. The dashed green line depicts the linear fit to the LASCO-C3 data³ (light green dots) for elongations greater than $13R_\odot$, extrapolated down to $4R_\odot$. The exponent, n , in this case is also

2.31 (note the match between the blue and green dashed lines). The LASCO-C3 data were normalized to the WISPR value at $20R_\odot$. For WISPR the absolute calibration was determined by analysing the intensity of stars in the field, which resulted in an error of 12%. The relative accuracy and repeatability of the WISPR are excellent, which gives us high confidence in the turnover of the intensities below $17R_\odot$. The historical measurements represented by the black dots²⁴ and green dots³ both have absolute errors of 20%. On the other hand, no error was given for the data represented by the red dots²⁵.

Figure 2 shows the K-corona from both telescopes on 6 November 2018, after removal of the brighter background from the dust scattering. Supplementary Videos 1 and 2 provide background-removed videos of the images taken during the first two encounters. The grid lines for both Fig. 2 and the Supplementary Videos are in the HPLN-ARC, HPLT-ARC coordinate system^{26,27}. The videos show the evolution of a coronal streamer during the two encounter periods of PSP observations. On the large scale, the agreement with model predictions

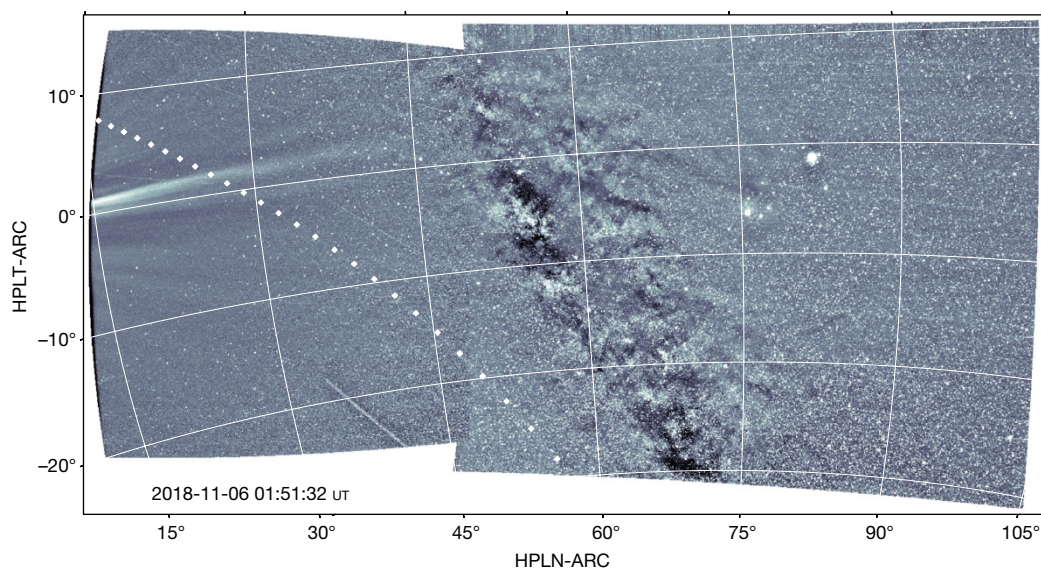


Fig. 2 | Combined images from the inner and outer telescopes of WISPR on 6 November 2018 at 01:44 UT. After removal of an empirical model of the F-corona, the faint solar wind structures are revealed. A faint streamer outlining the heliospheric current sheet is visible, as are faint, radial and diffuse rays, all with apparent origin on the Sun. The image also reveals the dust trail

along the orbit of the asteroid 3200 Phaethon (delineated by the white dots). The Galaxy dominates the scene in the inner part of the outer telescope accompanied by two bright objects: Jupiter (to the upper right) and the star Antares (a little below to its left) in the Scorpius constellation.

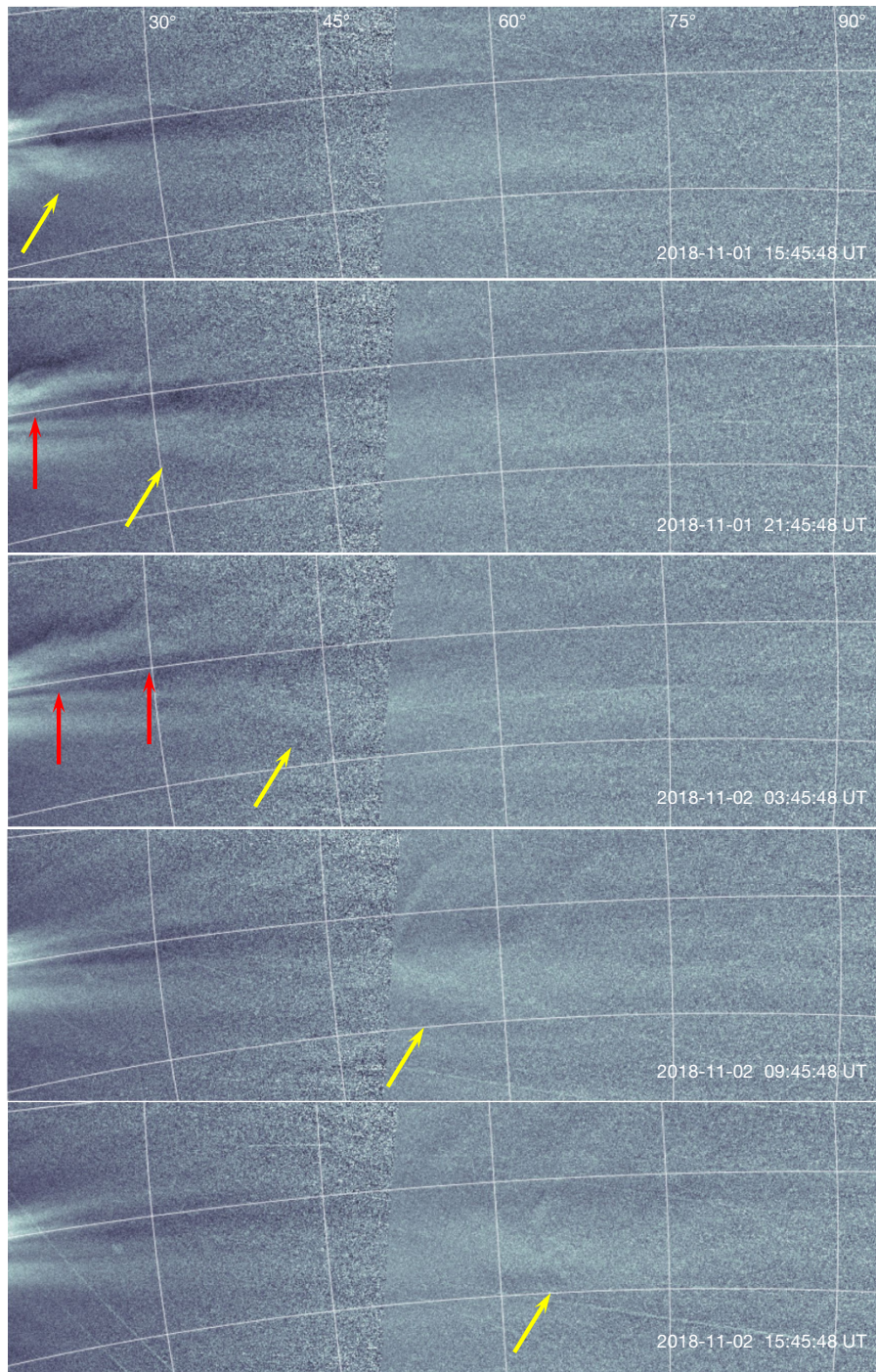


Fig. 3 | The propagation of a CME. Shown are five cropped frames from Supplementary Video 1 at different times in the same coordinate system as Fig. 2. The radial range is shown at the top, and the latitudinal range is $0^\circ \pm 10^\circ$ for each panel. The yellow and red arrows point to structures described in the text.

by our team (Extended Data Fig. 1) is very good, validating the model assumptions about the configuration of the magnetic field and the mass flux of the equatorial solar wind. The representation of WISPR-I images in a latitude versus time format (Extended Data Fig. 2a) reveals that near perihelion WISPR suddenly imaged faint coronal rays that are distinct from the main streamer rays. We note that fine structure along the streamer belt has been observed before^{19,28}. High-resolution simulations of the corona reproduce these brightness features. We interpret their displacement to higher apparent latitudes to the spacecraft motion (Extended Data Fig. 2b). This striated ‘texture’ of the background corona is caused in our model by the spatial variability

of coronal magnetic flux tubes, along which the plasma is heated and accelerated to form the slow solar wind.

Supplementary Video 1 shows a series of ejecta along the streamer. A particular event characterized by a big magnetic flux rope followed by several smaller ones is shown in Fig. 3. The first one (yellow arrows) has an elliptical high-density envelope surrounding a quasi-circular density depletion at its centre and a striated envelope. Although similar structures have been observed by LASCO, they could not always be resolved. The Encounter 1 images were binned 2×2 pixels, giving an effective 2-pixel spatial resolution¹⁵ of 60 arcsec (at 0.21 AU) for WISPR-I, which is about $2 \times$ finer than the LASCO-C3 observations of this event.

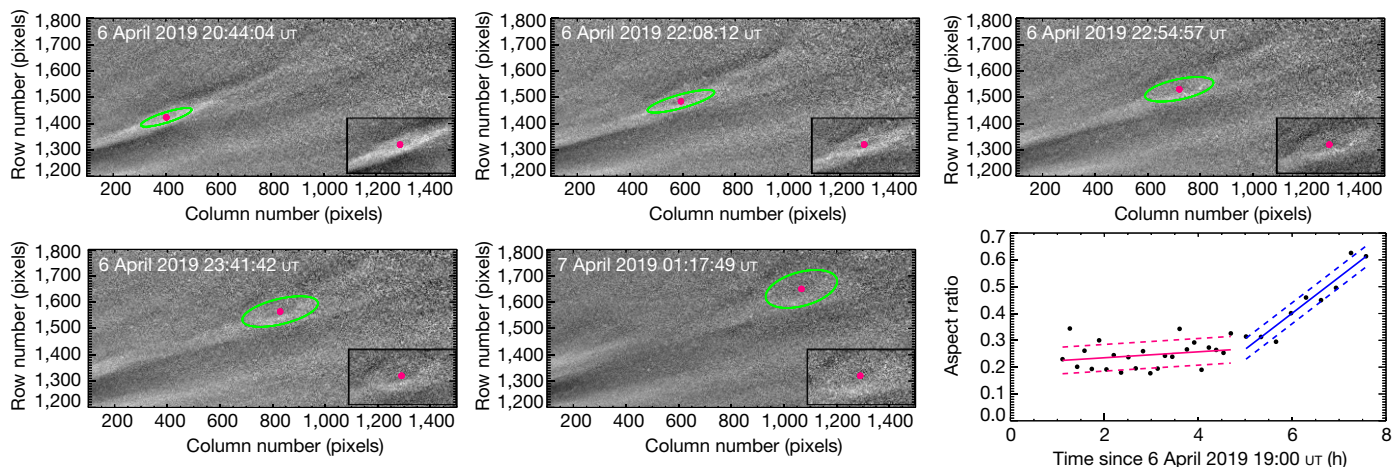


Fig. 4 | Formation and propagation of an island-like structure within a streamer. The first five panels show snapshots from WISPR-I during the second perihelion. An ellipse (green line) is fitted to the high-density ring of the structure in each panel. The final panel shows the aspect ratio (minor axis/

major axis) of these ellipse fits versus time. The red and blue solid lines show a linear fit to this aspect ratio from 20:00 UT to 23:45 UT on 6 April 2019 (red) and from 00:00 UT to 02:40 UT on 7 April 2019 (blue). The pairs of dashed lines on either side of these fits show the 1-sigma values.

On the other hand, LASCO-C2 tracked this structure when it was much closer to the Sun for a few images only, but with about 2.5× better spatial resolution. The event was also recorded by WISPR-O with an effective 2-pixel spatial resolution of 96 arcsec, extending the coverage of the event. Such density features have been interpreted as the boundaries of magnetic flux ropes^{12,16}. We have combined the spatially resolved density information with modelling to locate the structures corresponding to the internal toroidal and poloidal magnetic fields and study their interactions with the ambient plasma as the structure expands inside the streamer rays. Preliminary work demonstrates (Extended Data Fig. 3) that the structures are indeed consistent with a force-free magnetic flux rope propagating along the heliospheric current sheet, which is quite flat during this period. The heliospheric current sheet flatness may be the reason for detecting the fine-scale structure of the event. Such behaviour is extremely rare in 1 AU observations²⁹.

The event shown in Fig. 3 (on 1 and 2 November 2018) also shows two additional smaller flux ropes (red arrows) following the northern boundary of the main flux rope; these are probably by-products of the interaction between the main event and the ambient corona. The quasi-circular shape and faint striations within the feature strongly suggest that it is an idealized magnetic flux rope. Smaller features, with similar morphologies, are also seen following the main ejection. The yellow arrows follow the first, main event. The red arrows follow the two following events, until they become merged with the background, although they are still visible in Supplementary Video 1. These structures were not detected by either LASCO or SECCHI, although both have observed many small ejecta¹⁴. Small dense features caused by interaction of a coronal mass ejection (CME) with its environment have proved difficult to identify positively³⁰. Further studies will be necessary to test that idea.

Observations during the second perihelion (Supplementary Video 2) again show new dynamics in a coronal streamer. In this case, the observations capture the formation of oblong structures consistent with magnetic islands. Magnetic islands are, in two dimensions, a collection of roughly elliptical magnetic field lines that close on themselves; or, in three dimensions, helical field lines wrapping around a central (guide) field, again with a roughly elliptical cross-section. These island structures are predicted to form via the tearing-mode instability¹⁷ from magnetic reconnection in a current sheet, such as the one within this streamer, where oppositely directed magnetic fields meet. Figure 4 shows several snapshots of this streamer and the formation and evolution of one of these oblong structures. This structure first

appears at the inner edge of the image around 6 April 2019 20:00 UT and propagates out, within the current sheet, as an expanding, highly elliptical shape with a high-intensity (dense) ring of emission surrounding a low-intensity core. The final panel of Fig. 4 shows a measure of the aspect ratio (ratio of minor axis to major axis) of the ellipses fitted to this structure in each of the 33 frames from time 6 April 2019 19:57 UT to 7 April 2019 02:54 UT. Each ellipse was fitted to a set of points placed by hand on the high-density ring of the oblong structure in each frame. The corresponding ellipse is shown as a green curve in the five snapshots here, with a red dot at the centre of the ellipse. The plot indicates that the structure expands with a slightly increasing aspect ratio until 23:45 UT on 6 April 2019 and then it increases more quickly until the entire structure fades into the background. This evolution, including the increase in aspect ratio, is consistent with simulations of the tearing-mode formation of islands in an expanding coronal wind¹⁸. These simulations show the un-reconnected guide field collecting at the centre, forming this low-emission core, with the reconnecting field forming the high-density ring around the core. Although such an island ejection from a coronal streamer has been reported previously³¹, the earlier observations were not sufficiently resolved to show this internal ring and core structure.

WISPR imaged a variety of interesting structures in the corona/solar wind during the first two PSP orbits about the Sun. The departure from linearity of the F-corona intensity profiles below about 17R_o is opposite to that found in both Helios and STEREO data. Although this behaviour could be leading to the predicted dust-free zone close to the Sun, the intensity decrease could be due to a change in the properties of the dust scattering, or a combination of the two. WISPR has certainly not observed the dust-free zone. Theoretical analyses of the plausible existence of a dust-free zone predict^{8–11} the formation of circumsolar dust bands that could be observed by their thermal emission. In a compilation of the 30 observations⁶ made at various wavelengths from 0.8 μm to 3.6 μm during eleven solar eclipses from 1966 through to 1998, about half indicated an enhancement and the other half, including the two latest eclipses in 1991 and 1998, did not. The resolution of whether this WISPR finding represents dust depletion or something else will have to wait until PSP steps down to lower perihelia.

The near-corotation of PSP allows us to observe the radial outflow of the solar wind, without the confusing impact of solar rotation. The observations suggest that many small ejecta, commonly called ‘blobs’, may indeed be magnetic flux ropes but are usually too small to identify as such from 1 AU (ref. ³²). Structures larger than these are generally

interpreted as CMEs, but the physical mechanism of formation may not be the same. This finding, particularly with the anticipated measurements of the same structures by PSP's in situ payload, may finally clarify the evolution of the CME magnetic structure in the heliosphere, opening up avenues of research on internal CME dynamics. As PSP steps closer to the Sun over the next five years, these observations, together with the modelling, will certainly provide insights and opportunities to study and separate the spatial and temporal variability of the solar wind near its source and will probably increase the performance of space weather prediction schemes. This will benefit a wide range of communities from basic physics research to space situational awareness to even astrophysics through exoplanet habitability applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1807-x>.

- Koutchmy, S. & Lamy, P. L. The F-corona and the circum-solar dust evidences and properties. In *IAU Colloq. 85: Properties and Interactions of Interplanetary Dust* (eds Giese, R. H. & Lamy, P.) (Reidel, 1985).
- MacQueen, R. M. et al. The outer solar corona as observed from Skylab: preliminary results. *Astrophys. J.* **187**, L85–L88 (1974).
- Brueckner, G. et al. The large angle spectroscopic coronagraph (LASCO). *Sol. Phys.* **162**, 357–402 (1995).
- Howard, R. A. et al. Sun–Earth connection coronal and heliospheric investigation (SECCHI). *Space Sci. Rev.* **136**, 67–115 (2008).
- Leinert, C. & Grun, E. Interplanetary dust. In *Physics of the Inner Heliosphere Vol. I* (eds Schwenn, R. & Marsch, E.) (Springer, 1990).
- Mann, I. et al. Dust near the Sun. *Space Sci. Rev.* **110**, 269–305 (2004).
- Leinert, C., Hanner, M., Link, H. & Pitz, E. Search for a dust free zone around the Sun from the Helios 1 solar probe. *Astron. Astrophys.* **65**, 119–122 (1978).
- Lamy, P. et al. No evidence of a circumsolar dust ring from infrared observations of the 1991 solar eclipse. *Science* **257**, 1377–1380 (1992).
- Russell, H. N. On the composition of the Sun's atmosphere. *Astrophys. J.* **70**, 11 (1929).
- Lamy, P. L. The dynamics of circum-solar dust grains. *Astron. Astrophys.* **33**, 191–194 (1974).
- Mukai, T. & Yamamoto, T. On the circumsolar grain materials. *Publ. Astron. Soc. Jpn.* **26**, 445–458 (1979).
- Vourlidas, A., Lynch, B. J., Howard, R. A. & Li, Y. How many CMEs have flux ropes? Deciphering the signatures of shocks, flux ropes, and prominences in coronagraph observations of CMEs. *Sol. Phys.* **284**, 179–201 (2013).
- Fox, N. J. et al. The Solar Probe Plus mission: humanity's first visit to our star. *Space Sci. Rev.* **204**, 7–48 (2016).
- Sheeley, N. R. Jr, Lee, D. D.-H., Casto, K. P., Wang, Y.-M. & Rich, N. B. The structure of streamer blobs. *Astrophys. J.* **722**, 1522–1538 (2010).
- Vourlidas, A. et al. The Wide-Field Imager for Solar Probe Plus (WISPR). *Space Sci. Rev.* **204**, 83 (2016).
- Chen, J. et al. Magnetic geometry and dynamics of the fast coronal mass ejection of 1997 September 9. *Astrophys. J.* **533**, 481 (2000).
- Furth, H. P., Killeen, J. & Rosenbluth, M. N. Finite-resistivity instabilities of a sheet pinch. *Phys. Fluids* **6**, 459 (1963).
- Rappazzo, A. F., Velli, M., Einaudi, G. & Dahlburg, R. B. Diamagnetic and expansion effects on the observable properties of the slow solar wind in a coronal streamer. *Astrophys. J.* **633**, 474 (2005).
- DeForest, C. E. et al. The highly structured outer solar corona. *Astrophys. J.* **862**, 18 (2018).
- Leinert, C., Richter, I., Pitz, E. & Planck, B. The zodiacal light from 1.0 to 0.3 A.U. as observed by the HELIOS space probes. *Astron. Astrophys.* **103**, 177–188 (1981).
- Eyles, C. J. et al. The heliospheric imagers onboard the STEREO mission. *Sol. Phys.* **254**, 387–445 (2009).
- Kaiser, M. L. et al. The STEREO mission: an introduction. *Space Sci. Rev.* **136**, 5–16 (2008).
- Stenborg, G., Howard, R. A. & Stauffer, J. R. Characterization of the white-light brightness of the F-corona between 5° and 24° elongation. *Astrophys. J.* **862**, 168 (2018).
- Saito, K., Poland, A. I. & Munro, R. H. A study of the background corona near solar minimum. *Sol. Phys.* **55**, 121–134 (1977).
- Allen, C. W. *Astrophysical Quantities* (University of London, 1955).
- Thompson, W. T. Coordinate systems for solar image data. *Astron. Astrophys.* **449**, 791–803 (2006).
- Calabretta, M. R. & Greisen, E. W. Representations of celestial coordinates in FITS. *Astron. Astrophys.* **395**, 1077–1122 (2002).
- Thernisien, A. F. & Howard, R. A. Electron density modeling of a streamer using LASCO data of 2004 January and February. *Astrophys. J.* **642**, 523–532 (2006).
- Vourlidas, A. et al. Comprehensive analysis of coronal mass ejection mass and energy properties over a full solar cycle. *Astrophys. J.* **694**, 1471–1480 (2009).
- Vourlidas, A., Maia, D., Pick, M. & Howard, R. A. LASCO/Nancay observations of the CME on 20 April 1998: white light sources of type-II radio emission. In *Magnetic Fields and Solar Prominences* (ed. Wilson, A.) SP448, 1003 (European Space Agency, 1999).
- Ko, Y.-K. et al. Dynamical and physical properties of a post-coronal mass ejection current sheet. *Astrophys. J.* **594**, 1068–1084 (2003).
- Rouillard, A. P. et al. The solar origin of small interplanetary transients. *Astrophys. J.* **734**, 7 (2011).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

WISPR contains two telescopes and measures the intensity of the visible light corona in addition to stellar and galactic sources. The two telescopes slightly overlap and have a combined field-of-view of 13.5° to 108.5° from the Sun, corresponding to approximately 9 to 78 solar radii (R_\odot ; $1R_\odot = 696,000$ km) at $35.6R_\odot$ perihelion. The visible light corona consists of two components: light scattered by free electrons (the K-corona) and light scattered by interplanetary dust (the F-corona). The F-corona of each WISPR image is removed using a technique³⁶ similar to that developed for the SECCHI/HI-1. The primary difference is that for the HI-1 images the initial step in the procedure analysed the horizontal lines in the image, whereas here, the initial step uses the vertical lines in the images.

All of the data presented here have been calibrated into Mean Solar Brightness (MSB) units. The calibration details will be published in a future paper, but include the removal of geometric distortion, vignetting, instrumental artefacts (stray light, and so on) and then applying the photometric calibration of the system. The vignetting is caused by two sources: the projection of the image onto the two-dimensional plane of the Advanced Pixel Sensor detector and for WISPR-I the obscuration of the objective lens of the sunward side of the image by a series of baffles (including the PSP heat shield) which are used to block the solar disk illumination and block diffraction from the edges of the preceding baffles. The absolute calibration is confirmed on-orbit by measuring the intensity of stars passing through the field. The intensity of the stars as they transit across the image is also a check on the vignetting correction.

Code availability

The code used in the WISPR pipeline and analysis is available as part of the SolarSoft library (<https://sohowww.nascom.nasa.gov/data/software.html>).

Data availability

The PSP Science Data Management Plan (https://sppgway.jhuapl.edu/docs/data/7434-9101_Rev_A.pdf) requires that all science data from the first two orbits with calibrations must be released to the public within six months of downlink of the first orbit. In addition to this data type, we will be releasing background subtracted images, videos, and lists

of events. Furthermore, the data must be delivered to the appropriate NASA/GSFC facility and integrated into the Virtual Observatory. Thus, the data is available from 12 November 2019. A complete archive is maintained at NRL (<https://wispr.nrl.navy.mil>) and will be publicly available at least during the full mission lifetime. A copy of the WISPR data will be located at the NASA/GSFC SDAC facility (<https://umbra.nascom.nasa.gov>) and integrated into the Virtual Solar Observatory.

33. Pinto, R. & Rouillard, A. P. A multiple flux-tube solar wind model. *Astrophys. J.* **838**, 89 (2017).
34. van Leeuwen, F. Validation of the new Hipparcos reduction. *Astron. Astrophys.* **474**, 653–664 (2007).
35. Arge, C. N. et al. Air force data assimilative photospheric flux transport (ADAPT) model. In *Twelfth International Solar Wind Conference* (eds Maksimovic, M., Issautier, K., Meyer-Vernet, N., Moncuquet, M. & Pantellini, F.) 343–346 (AIP, 2016).
36. Stenborg, G. & Howard, R. A. A heuristic approach to remove the background intensity on white-light solar images. I. STEREO/HI-1 heliospheric images. *Astrophys. J.* **839**, 68 (2017).

Acknowledgements We acknowledge the efforts of the PSP operations team in operating the mission and the WISPR team in developing and operating the instrument. We are grateful to R. Pinto (IRAP) for providing the Multi-VP magnetohydrodynamics simulations of the background solar wind used in Extended Data Fig. 3. R.A.H., A.V., R.C.C., C.E.D.F., B.G., J.R.H., P.H., A.K.H., C.M.K., P.C.L., J.L., M.L., N.E.R., D.G.S., G.S. and A.F.T. acknowledge support from the NASA Parker Solar Probe Program Office. N.M.V. is supported through the NASA Heliophysics Internal Scientist Funding Model. A.P.R., A.K. and N.P. acknowledge financial support from the ERC for the project SLOW_SOURCE - DLV-819189. P.L.L. acknowledges financial support from Centre National d'Etudes Spatiales. P.R. acknowledges support by the BELSPO /PRODEX. V.B. acknowledges the support of the Coronagraphic German and US Solar Probe Plus Survey (CGAUSS) project for WISPR by the German Aerospace Center (DLR) under grant 50OL1901 as a national contribution to the Parker Solar Probe mission.

Author contributions All authors contributed to writing the manuscript. R.A.H., A.V., N.R. and G.S. designed and collected the data. N.R., P.H., R.C.C., B.G. and G.S. performed the data processing and calibration. G.S. developed the technique for computing the background models. R.A.H., A.F.T., G.S. and P.L.L. performed the analysis of the dust scattering. A.V., C.E.D.F., M.L., P.H., P.C.L., A.R., N.P., A.K., N.V., G.S., A.K.H., N.E.R., V.B., P.R. and R.A.H. carried out the analysis of the K-corona. J.L. assisted in the observation planning by providing magnetohydrodynamics model predictions. A.K.H. and N.E.R. coordinated the data acquisition and downlink. P.C.L., J.R.H. and P.P. assisted with data calibration, observation planning and analysis. R.A.H., N.R., A.V., P.L.L., S.P.P., C.M.K., R.C.C. and D.G.S. assisted with design, calibration and instrument checkouts.

Competing interests The authors declare no competing interests.

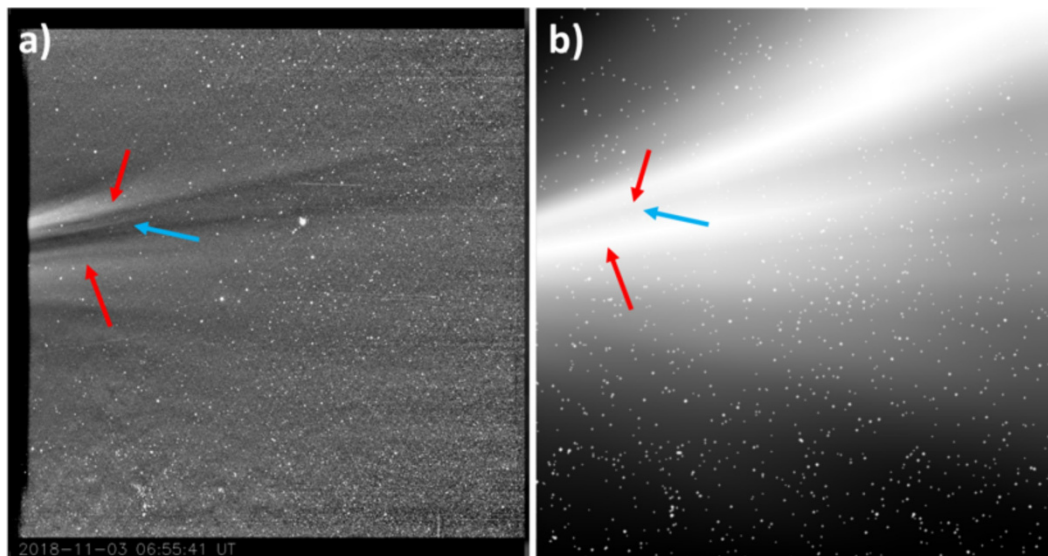
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1807-x>.

Correspondence and requests for materials should be addressed to R.A.H.

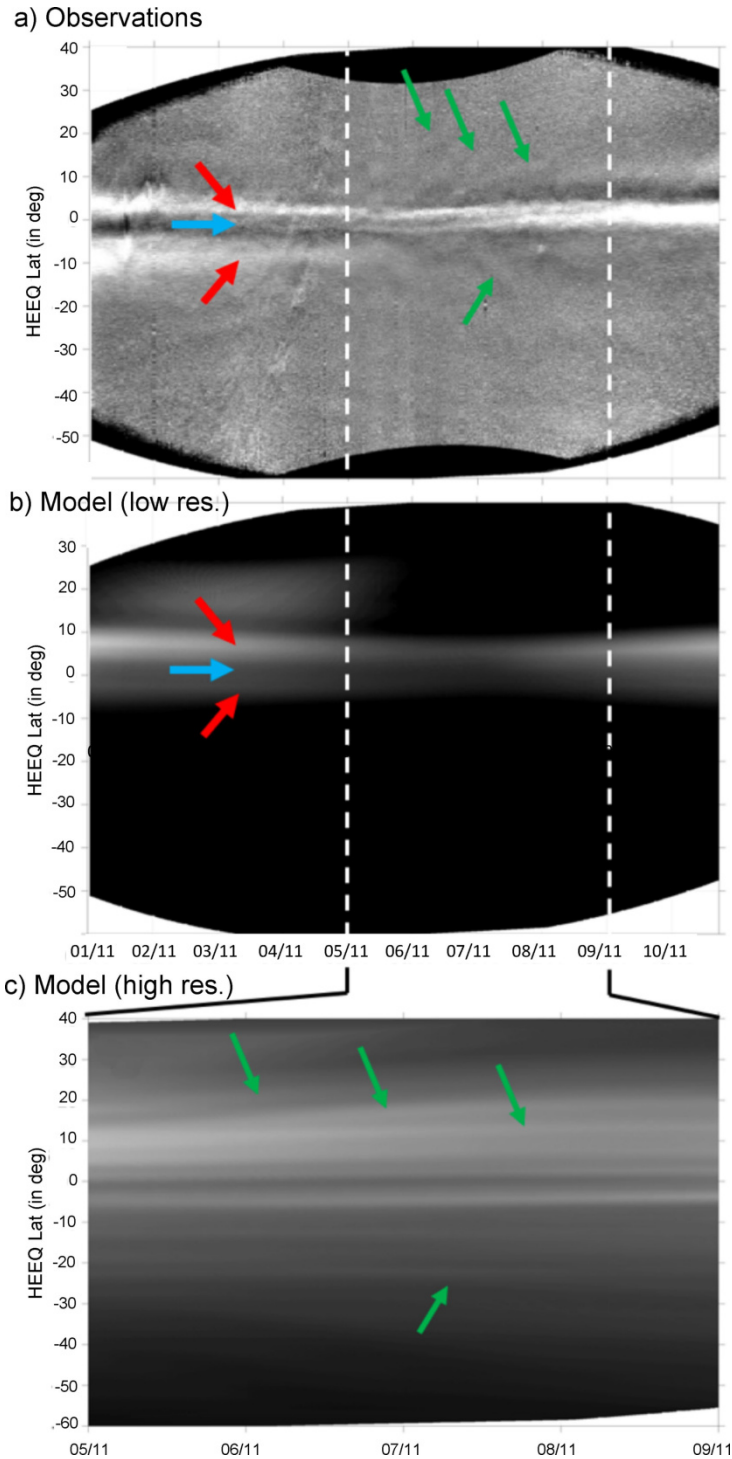
Peer review information *Nature* thanks Manuela Temmer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



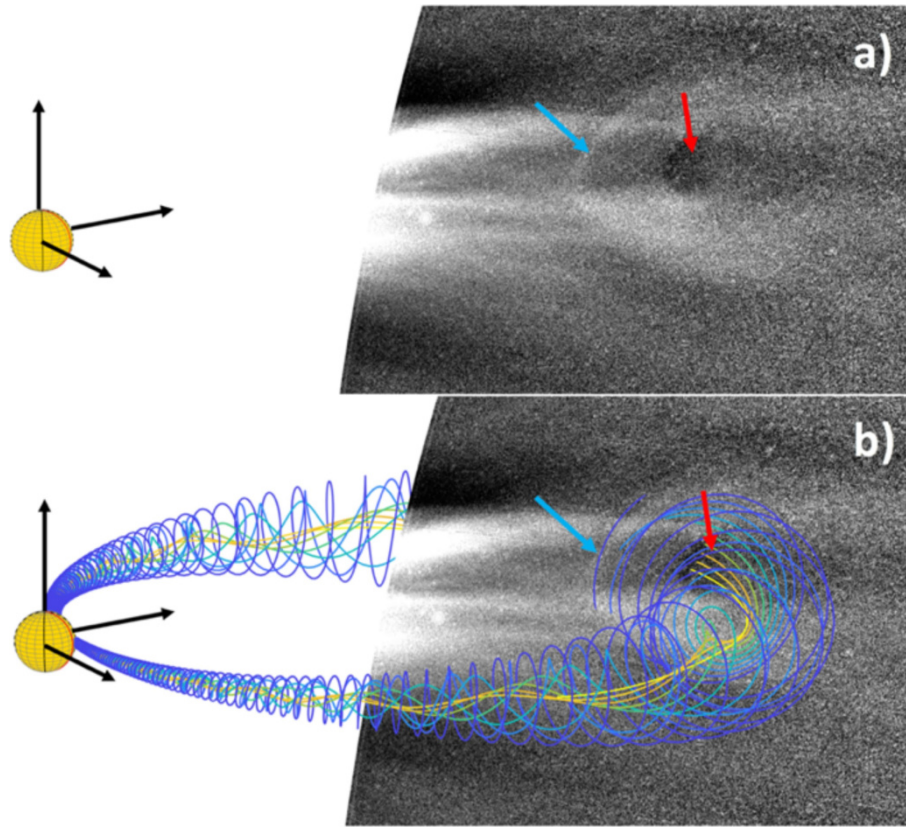
Extended Data Fig. 1 | Comparison of observations and synthetic observations from magnetohydrodynamics model. a, An image from the inner WISPR telescope taken on 3 November 2018 at 06:55:41 UT. The field of view (of both panels) is $40^\circ \times 40^\circ$ with the Sun 13.5° to the left. Two distinct sets of bright streamer rays are marked by red arrows. They are separated by a darker region marked by a blue arrow. The technique employed to remove the background F-corona in the WISPR image has artificially enhanced this dark region. The streamer rays located northwards of the dark region (top red arrow) are brighter than the rays situated southwards of the dark region (bottom red arrow). **b,** A synthetic white-light image produced from

three-dimensional simulations of the solar wind by the MULTI-VP magnetohydrodynamics code using a Wilcox Solar Observatory photospheric magnetogram³³. The three-dimensional density cubes produced by running the MULTI-VP code were processed by a white-light rendering code computing the brightness of the corona in the WISPR field of view from the heliocentric position of Parker Solar Probe. The MULTI-VP numerical model and the procedure to produce white-light images have been detailed³³. The star field from the new Hipparcus astrometric catalogue³⁴ was added to the simulated image in **b** for comparison with the WISPR image in **a**.



Extended Data Fig. 2 | Latitude versus time maps—observations and modelling. HEEQ, Heliocentric Earth Equator. **a**, A representation of WISPR inner telescope images in the form of a latitude versus time map. This map provides a summary of the temporal and spatial variability of coronal rays observed during the first encounter. We note that such fine structure along the streamer belt has been observed before^{19,28}. We identify in these maps the main streamer rays already seen in Extended Data Fig. 1 (the same blue and red arrows are shown here). During the period of super and corotation (5 to 9 November 2018), bright coronal rays drift in latitude away from the equator (green arrows). This is also visible in Supplementary Video 2. **b**, An equivalent map to **a** obtained from the WISPR synthetic images based on the MULTI-VP three-dimensional density cubes shown in Extended Data Fig. 1b. These medium-resolution simulations reproduce the time-varying aspect of the main streamer including their fading during perihelion (5 to 7 November).

c, MULTI-VP high-resolution simulation results for the period 5 to 9 November 2018 based on 2-degree resolution magnetograms produced by the Air Force Data Assimilative Photospheric Flux (ADAPT) model³⁵. The colour table has been saturated in these maps to enhance the features. The solar wind simulations reveal the finer striated structure of the corona and the coronal rays migrating poleward as observed by WISPR (green arrows). A search in the simulation data cubes reveals that these faint rays are separate from the brighter streamer rays. They form in the simulation as a result of considerable variability in the properties of the magnetic fields along which the slow solar wind forms. Since the prescribed coronal heating is scaled to the magnetic field properties this drives different mass flux along different flux tubes. We interpret the coronal rays marked by the top red arrows as resulting from the main streamer and the rays situated southwards (bottom red arrow) as resulting of a pseudo-streamer.



Extended Data Fig. 3 | Modelling of a CME as a 3D flux rope. a, An image from the inner WISPR telescope taken on 1 November 2018 at 19:30:50 UT during the passage of a pristine CME. Clear substructures are discernible in the WISPR image. The field of view is $40^\circ \times 40^\circ$ with the Sun 13.5° to the left. A bright ring at the outer contour/boundary of the CME is indicated by a blue arrow. A striking feature of this CME event is the presence of a dark circular core located at the centre of the CME event and indicated by a red arrow. **b,** The same image as in **a** but with the results of a three-dimensional flux rope fit superimposed. This figure proposes an interpretation for the different features observed by WISPR based on our current understanding of the appearance of CMEs imaged

in white light. The magnetic field lines (computed from solutions of the Grad-Shafranov equation) of the CME are traced inside this flux rope. The bright ring (blue arrow) corresponds to plasma located on the boundary of the flux rope where the poloidal magnetic field lines of the CME are adjacent to the ambient solar wind plasma. The dark core (red arrow) marks the location where strong toroidal (axial) magnetic fields dominate the plasma locally. Detailed modelling of the event will be presented in a future dedicated publication. We acknowledge the use of the Wilcox Solar Magnetograms used in this paper, obtained from the website at <http://wso.stanford.edu>.

Highly structured slow solar wind emerging from an equatorial coronal hole

<https://doi.org/10.1038/s41586-019-1818-7>

Received: 16 July 2019

Accepted: 11 November 2019

Published online: 4 December 2019

S. D. Bale^{1,2,3,4*}, S. T. Badman^{1,2}, J. W. Bonnell¹, T. A. Bowen¹, D. Burgess⁴, A. W. Case⁵, C. A. Cattell⁶, B. D. G. Chandran^{7,8}, C. C. Chaston¹, C. H. K. Chen⁴, J. F. Drake^{9,10,11}, T. Dudok de Wit¹², J. P. Eastwood³, R. E. Ergun¹³, W. M. Farrell¹⁴, C. Fong^{1,2}, K. Goetz⁶, M. Goldstein^{15,16}, K. A. Goodrich¹, P. R. Harvey¹, T. S. Horbury³, G. G. Howes¹⁷, J. C. Kasper^{5,18}, P. J. Kellogg⁶, J. A. Klimchuk¹⁹, K. E. Korreck⁵, V. V. Krasnoselskikh¹², S. Krucker^{1,20}, R. Laker³, D. E. Larson¹, R. J. MacDowall¹⁴, M. Maksimovic²¹, D. M. Malaspina¹³, J. Martinez-Oliveros¹, D. J. McComas²², N. Meyer-Vernet²¹, M. Moncuquet²¹, F. S. Mozer¹, T. D. Phan¹, M. Pulupa¹, N. E. Raouafi²³, C. Salem¹, D. Stansby³, M. Stevens⁵, A. Szabo¹⁹, M. Velli²⁴, T. Woolley³ & J. R. Wygant⁶

During the solar minimum, when the Sun is at its least active, the solar wind^{1,2} is observed at high latitudes as a predominantly fast (more than 500 kilometres per second), highly Alfvénic rarefied stream of plasma originating from deep within coronal holes. Closer to the ecliptic plane, the solar wind is interspersed with a more variable slow wind³ of less than 500 kilometres per second. The precise origins of the slow wind streams are less certain⁴; theories and observations suggest that they may originate at the tips of helmet streamers^{5,6}, from interchange reconnection near coronal hole boundaries^{7,8}, or within coronal holes with highly diverging magnetic fields^{9,10}. The heating mechanism required to drive the solar wind is also unresolved, although candidate mechanisms include Alfvén-wave turbulence^{11,12}, heating by reconnection in nanoflares¹³, ion cyclotron wave heating¹⁴ and acceleration by thermal gradients¹. At a distance of one astronomical unit, the wind is mixed and evolved, and therefore much of the diagnostic structure of these sources and processes has been lost. Here we present observations from the Parker Solar Probe¹⁵ at 36 to 54 solar radii that show evidence of slow Alfvénic solar wind emerging from a small equatorial coronal hole. The measured magnetic field exhibits patches of large, intermittent reversals that are associated with jets of plasma and enhanced Poynting flux and that are interspersed in a smoother and less turbulent flow with a near-radial magnetic field. Furthermore, plasma-wave measurements suggest the existence of electron and ion velocity-space micro-instabilities^{10,16} that are associated with plasma heating and thermalization processes. Our measurements suggest that there is an impulsive mechanism associated with solar-wind energization and that micro-instabilities play a part in heating, and we provide evidence that low-latitude coronal holes are a key source of the slow solar wind.

The first solar encounter of the Parker Solar Probe (PSP) occurred during the solar minimum. The spacecraft orbit remained within 5° of the heliographic solar equator and, unlike any previous spacecraft, was corotational with the Sun for two intervals surrounding perihelion. Figure 1 summarizes the radial magnetic field (B_r ; in heliocentric RTN coordinates; see Methods) structure observed by the FIELDS experiment¹⁷ for a six-week time interval centred on perihelion (6 November 2018). Figure 1a shows 1-s cadence measurements of B_r (see Methods) which show the overall $1/r^2$ behaviour expected from simple flux-conservation arguments¹⁸ as the heliocentric distance of PSP varied along its eccentric orbit. Against this background, dramatic and unexpected rapid polarity reversals where $\delta B_r/|B|$ is of the order of 1 are superposed ($|B|$ is the magnitude of the magnetic field). One-hour statistical modes

(most probable value; see Methods) of B_r in Fig. 1b remove the transient polarity inversions and reveal the large-scale magnetic structure. Time series predictions of B_r generated from the simple, but widely used, Potential Field Source Surface (PFSS) model^{19–21} are shown for comparison in black and green. The implementation of this model and the procedure to connect it to the location of PSP and generate the time series is discussed in the Methods section ‘PFSS modelling and connection to PSP’.

Magnetic field structure

PFSS is a zero-current force-free model of the global solar corona, meaning that it assumes that magnetic pressure dominates over gas

The list of affiliations appears at the end of the paper.

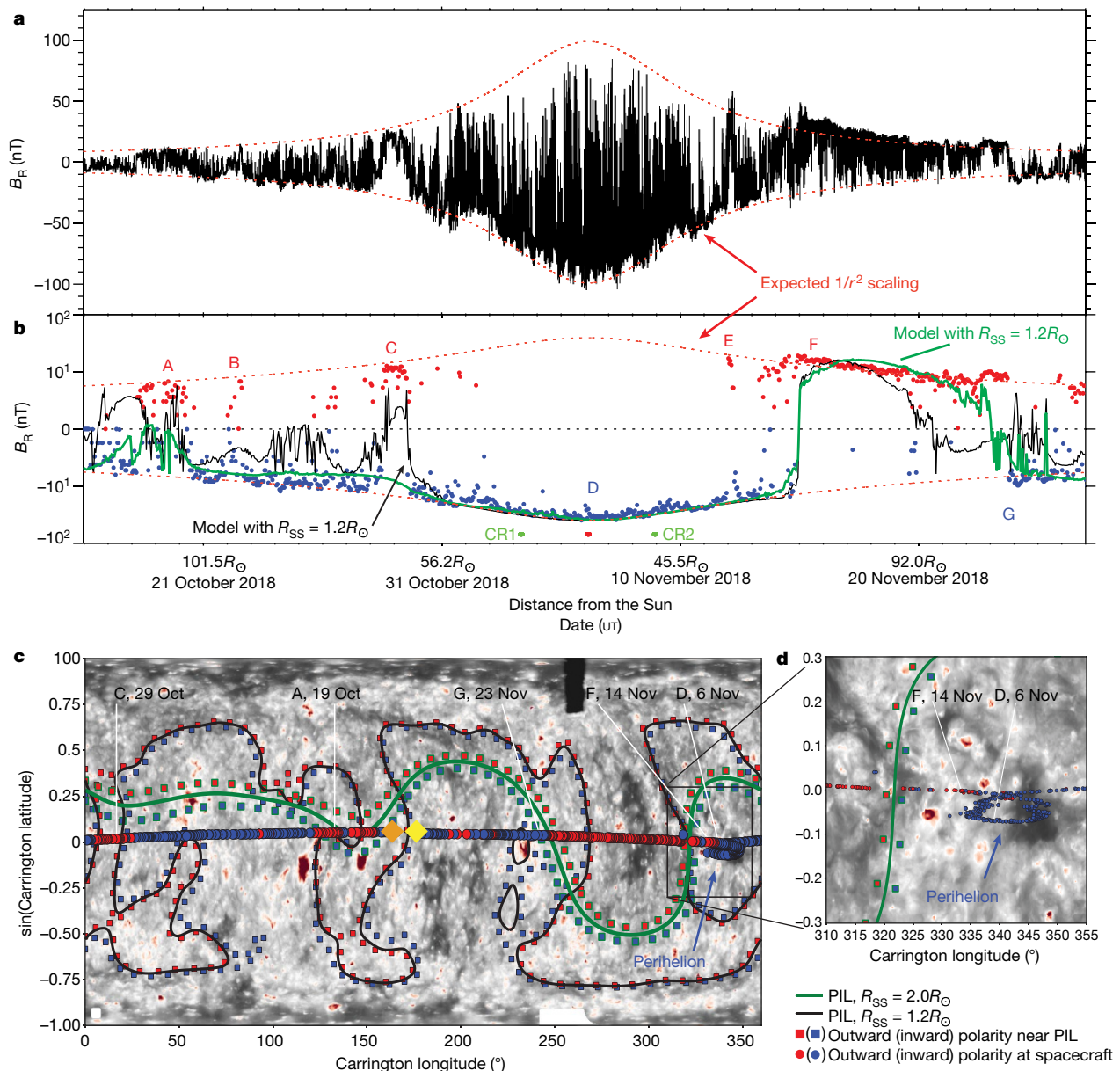


Fig. 1 | Radial magnetic field measurements are highly structured, map back to the Sun, and are consistent with a low source surface. a, The measured radial magnetic field B_R is comprised of a large-scale field, which scales approximately as $1/r^2$ (red dotted lines; also in Fig. 1b) and rapid large-amplitude polarity reversals ($\delta B_R/|B|$) of the order of 1 associated with jets of plasma (Fig. 2b). **b,** One-hour statistical modes of B_R (bisymmetric logarithmic plot) show the large-scale radial field coloured for polarity (red, outward; blue, inward). Predicted radial-field profiles from a PFSS model are over-plotted using a source-surface radius of $R_{SS} = 1.2R_\odot$ (black curve, unscaled) and $2.0R_\odot$ (green curve, multiplied by a factor of 6.5). R_{SS} at $1.2R_\odot$ reproduces many of the measured polarity changes (labelled A, C, F and G). The $R_{SS} = 2.0R_\odot$ model better predicts the timing of polarity inversion G (see Methods). Labels B and E indicate transient events, and the perihelion coronal hole interval is centred on D. Corotations (CR1 and CR2; green) and the perihelion (PH; red dot) at $35.7R_\odot$ are labelled. **c,** An extreme-ultraviolet synoptic map of 171-Å (Fe IX) emission

shows structure associated with active regions (small-scale extreme-ultraviolet bright points appear as red patches) and lower-density plasma in coronal holes (darker regions). The PSP trajectory at the source surface is superimposed, coloured as in **b** for measured field polarity. Encounter 1 begins at the orange diamond, moves westwards (in decreasing Carrington longitude, with respect to a fixed point on the solar surface) across the map through perihelion at about 330° , and ends at the yellow diamond. A line shows the location of the model polarity-inversion line (PIL) at the source surface ($R_{SS} = 1.2R_\odot$, black; $R_{SS} = 2.0R_\odot$, green). Red and blue squares indicate the polarity on either side of the PIL models. Red ($B_R > 0$) and blue ($B_R < 0$) lines map the magnetic field from R_{SS} back to the photosphere for $R_{SS} = 2.0R_\odot$; for $R_{SS} = 1.2R_\odot$ the model field lines are radial. **d,** The extreme-ultraviolet map of the perihelion interval, showing field lines mapping back to the Sun into a small equatorial coronal hole, and the location of the adjacent PIL associated with the heliospheric current sheet, from the $2.0R_\odot$ model.

pressure (that is, it assumes a low plasma β value), so that the problem reduces to magnetostatics, giving a solution of a static field configuration that rigidly corotates with the Sun. The role of gas dynamics is approximated by requiring that the tangential magnetic field vanishes at a spherical ‘source surface’ at some radius R_{SS} , which simulates how the outflowing solar wind drags the field lines out into the heliosphere.

The magnetostatic approximation limits the accuracy and applicability of the model. Nevertheless, PFSS is widely used as a computationally tractable first approximation and forms the basis for more sophisticated models^{21,22}. We note that PSP encounter 1 took place very close to solar minimum, with low solar activity, reducing the impact of non-potential transient events and active regions.

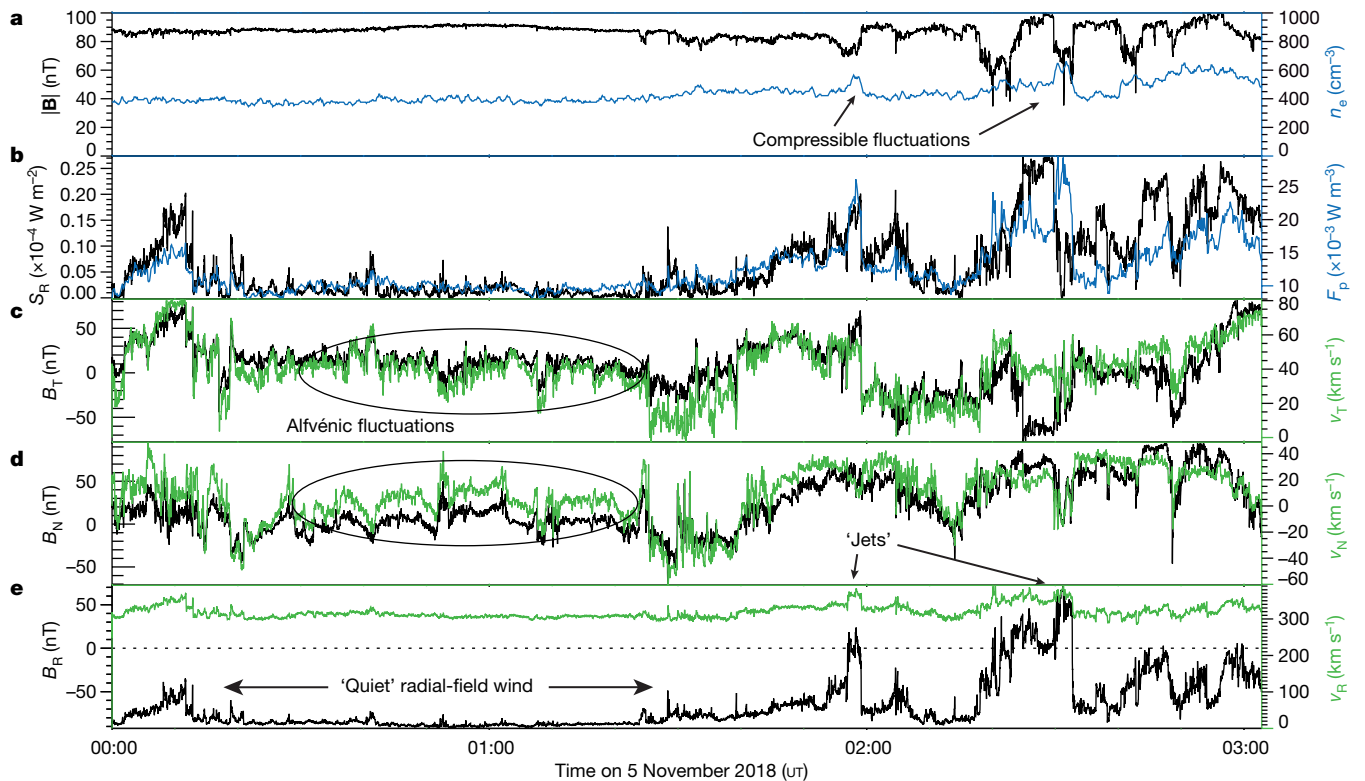


Fig. 2 | Magnetic field reversals and plasma jets carry a Poynting flux. **a**, Time series measurements of the magnetic field magnitude $|B|$ (black line) and total plasma density n_e (blue line) show anti-correlation during jet events, consistent with magnetohydrodynamic slow-mode behaviour. **b**, Radial Poynting flux (S_R , black line) and ion kinetic energy flux (F_p , blue line), showing large enhancements during jet-field reversal events. **c**, Tangential components of the magnetic field (B_T ; black) and plasma velocity (v_T , green) components

showing Alfvénic fluctuations. **d**, The normal components of the magnetic field (B_N , black line) and the plasma velocity (v_N , green line). **e**, Radial magnetic field (B_R , black line) and plasma velocity (v_R , green line), showing an interval of quiet radial-field wind and flow adjacent to a magnetic structure associated with jets of plasma. Measurements are made from around 00:00 to 03:00 5 November 2018 universal time (UT) at about $36.6R_\odot$. The Alfvén speed during the quiet interval is $v_A \approx 100 \text{ km s}^{-1}$.

In Fig. 1b, two model evaluations are shown with $R_{SS} = 2.0R_\odot$ (R_\odot , solar radius; green line) and $R_{SS} = 1.2R_\odot$ (black line). In both cases R_{SS} is well below the established value²³. However, this value is necessary to provide good agreement for all model inputs (see Methods) and is not without precedent^{24,25}. Model comparison reveals an overall very good agreement for both models, but also shows that the polarity inversions at features A and C are washed out except with the lower source-surface radius (black line). Meanwhile, the timing of feature G is better captured with the higher source-surface radius (green line), illustrating the difficulty that PFSS has with assuming a single source-surface radius, and supports previous findings of a varying ‘true’ source-surface radius^{25,26}. Finally, Fig. 1c, d depicts field-line mappings derived from the same PFSS models shown in Fig. 1b to connect the spacecraft down to the lower corona to establish context for the in situ measurements. The spacecraft trajectory is shown projected onto the source surface, coloured by its measured polarity.

The background is a synoptic map of extreme-ultraviolet emission in the 171-Å wavelength for which dark regions imply lower-density plasma and the likely location of open magnetic field lines. For reference, this background is shown in isolation in Extended Data Fig. 4 along with its corresponding map for the 193-Å wavelength. The neutral lines derived from the PFSS models are shown as single contours in the same colour as their time series in Fig. 1b. Figure 1c shows how the neutral line topology explains the polarity inversions measured by PSP. Figure 1d is a magnification of the two-week interval closest to perihelion (330° longitude). During the entire two-week corotation loop period, PSP remained connected to a small, negative-polarity, isolated equatorial coronal hole, suggesting that the rapid magnetic field polarity reversals seen in Fig. 1a are magnetic structures emerging

from this coronal hole and sweeping past the PSP spacecraft. Extended Data Fig. 5 shows the configuration schematically. For most of this interval, SWEAP²⁷ measurements of the solar-wind velocity indicated an Alfvénic slow wind stream (see Fig. 2), suggesting a slow solar wind source rooted in an equatorial coronal hole at the Sun. Polarity inversions B and E in Fig. 1b are associated with (transient) flux rope and coronal mass ejection²⁸ events, respectively.

Alfvénic fluctuations and plasma jets

Time series magnetic field and velocity structures show the correlations (Fig. 2c–e) expected of propagating Alfvén waves²⁹, especially during the quiet radial-field intervals. The δB_R polarity reversal intervals show enhanced radial wind velocity v_R (Fig. 2e) and the Alfvénic correlations of velocity and magnetic field (δv to δB) within the polarity inversions and jets suggest that these structures may be interpreted as large-amplitude, three-dimensional Alfvénic structures convected away from the Sun. As a simple measure, statistics of zero-crossings (polarity reversals; see Methods) show that around 6% of the temporal duration of encounter 1 is comprised of these so-defined jets. Many jet intervals show signatures of compressibility (Fig. 2a)—in this case anti-correlated plasma density n_e and $|B|$ —suggesting slow-mode or pressure-balanced behaviour³⁰. Isolated Alfvénic features associated with magnetic field reversals have been identified at $60R_\odot$ (ref. ³¹), near one astronomical unit (AU)³² and in the polar heliosphere by the spacecraft Ulysses³³; however, at those distances few or no compressive signatures were present. It has been suggested³⁴ that these magnetic structures could be signatures of impulsive reconnection events in the Sun’s atmosphere³⁵; simulations³⁶ show qualitative similarities to the events of encounter 1 but do not reproduce the observed magnetic field reversals past 90° .

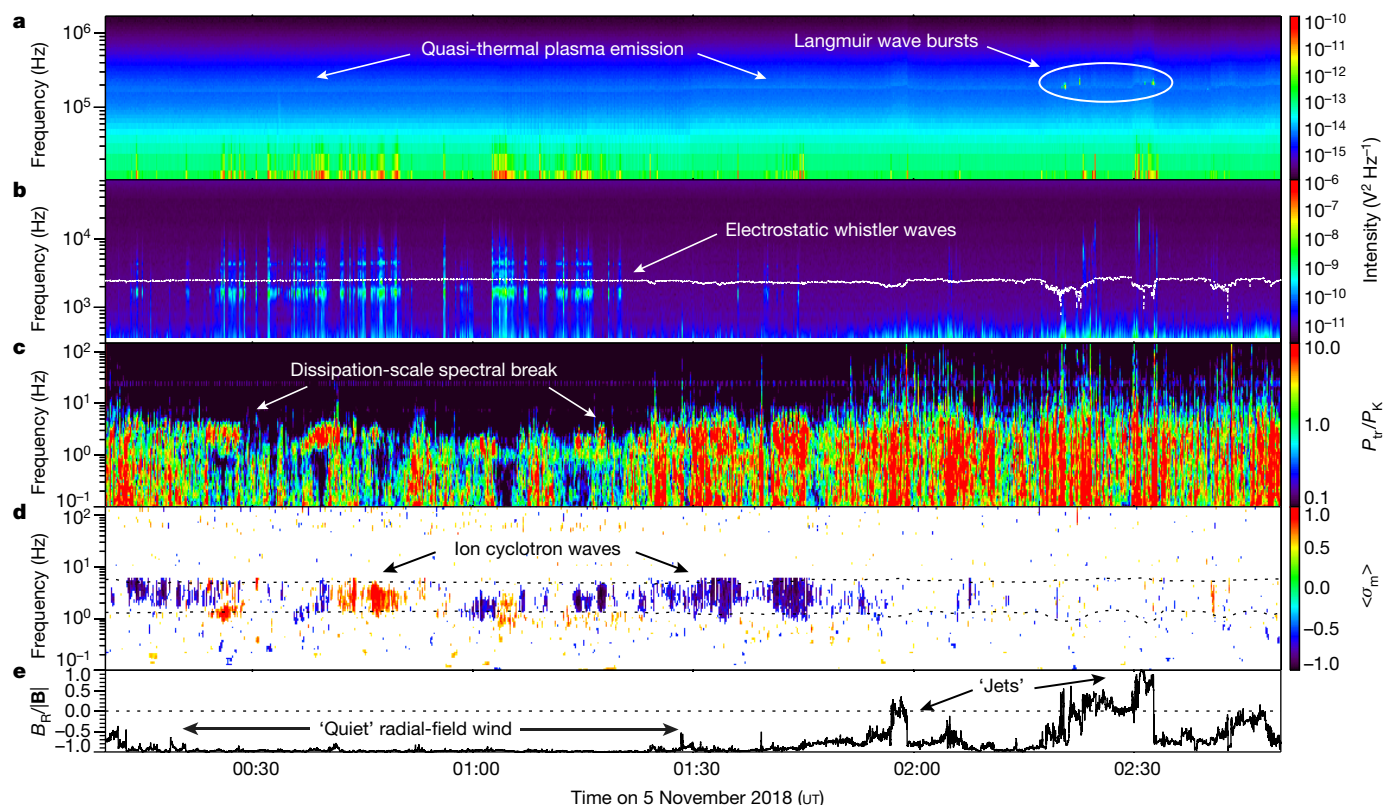


Fig. 3 | Plasma-wave activity near the perihelion differs in quiet wind and jets. a, Spectral density measurements of electric field fluctuations near the electron plasma frequency f_{pe} show intense bursts of electrostatic Langmuir waves with intensities around 10^2 – 10^4 V² Hz⁻¹ above the thermal background, suggesting the presence of electron beams. **b**, Electrostatic waves near the electron cyclotron frequency f_{ce} (white dotted line) and its harmonics are often present in intervals of ambient radial magnetic field, but not present in jet plasma. **c**, A wavelet spectrogram of the magnetic field shows bursts of turbulent fluctuations with a distinct spectral break between 1 and 10 Hz associated with a transition to dissipation scales. The trace magnetic

field intensity (P_{tr}) is divided by a factor of $P_k \propto f^{-5/3}$ such that a power spectrum with index $-5/3$ has no frequency dependence. **d**, Magnetic helicity (σ_m , from the wavelet spectrogram) shows narrowband $f_{ci} < f < f_{ci} + v_R/v_A$ (the expected Doppler-shifted frequency, where v_A is the local Alfvén speed; dashed lines) signatures associated with ion cyclotron waves—when σ_m is close to the maximum (+1) or the minimum (−1)—again in quiet radial solar wind. **e**, The normalized radial magnetic field $B_R/|B|$ shows distinct intervals of quiet radial-field wind, reduced turbulent levels and increased occurrence of electrostatic whistler and ion cyclotron instability. Measurements in Fig. 3 were made at approximately 00:00–03:00 November 5 2018 UT at about $36.6R_\odot$.

Micro-instabilities and turbulence

Alfvénic structures and waves have long been considered to be an important energy source for the solar wind^{11,12}. The radial Poynting flux $S_R = (\mathbf{E} \times \mathbf{B})_R/\mu_0$ (where μ_0 is the vacuum permeability and \mathbf{E} is the electric field; see Methods) in the spacecraft frame (Fig. 2b) is about 10% of the kinetic energy flux (blue curve) and shows enhancements during the jet intervals, suggesting that these plasma jets may impart energy to the emerging solar wind. As seen in Figs. 1a, 2e, the plasma jets appear to be clustered and interspersed in an otherwise quiet solar wind flow with prominently radial magnetic field.

The quiet radial-flow intervals contain plasma waves consistent with expectations of micro-instabilities associated with ion¹⁴ and electron¹⁶ velocity-space structure (Fig. 3). The electric field spectrum from about 11 kHz to about 1,688 kHz shows signatures of plasma quasi-thermal noise³⁷ (Fig. 3a) at the electron plasma frequency f_{pe} (which is used to estimate the total plasma density in Fig. 2a). Intense bursts of narrowband electrostatic Langmuir waves (Fig. 3a) occur throughout the perihelion encounter; narrowband Langmuir waves are driven by electron beams and damp rapidly, suggesting the presence of an intermittent, local population of electron beams.

The electric field spectrum (Fig. 3b) from 0.3 kHz to about 75 kHz shows intermittent bursts of electrostatic whistler-wave activity, peaked in power below the electron gyrofrequency f_{ce} . Also present are waves containing harmonic structure consistent with electron-Bernstein-wave emission. Electrostatic whistler–Bernstein bursts¹⁶

are generated by features in the electron velocity distribution function $f_e(v)$ and are not observed in the solar wind at 1 AU. Here they occur only in the quiet radial-field intervals. A wavelet spectrogram (divided by a function $P_k \propto f^{-5/3}$) of the search coil magnetometer and fluxgate magnetometer data in Fig. 3c shows the spectral content of the magnetic field to around 146 Hz. A spectral break between 1–10 Hz (in the spacecraft frame) is highly variable and associated with the transition from a magnetohydrodynamic turbulent cascade to dissipation or dispersion ranges at ion kinetic lengthscales³⁸. Note that the overall turbulent levels are lower and more intermittent in the quiet radial wind (Figs. 3c, 4a). The spectrum of magnetic helicity³⁹ σ_m in Fig. 3d indicates intervals of large ($1 > \sigma_m > 0.5$, red; $-0.5 < \sigma_m < -1$, blue) circular polarization often associated with ion cyclotron waves⁴⁰. These ion wave events are apparent during quiet radial-field intervals.

The (trace) magnetic field spectra (see Methods), averaged over 30 min (Fig. 4a), show broken power-law behaviour, with spectral indices roughly comparable to the $-5/3$ and $-8/3$ predictions for magnetohydrodynamic and kinetic-scale turbulence³⁸, respectively. This suggests that by $36.6R_\odot$, the solar wind has already developed a turbulent cascade, to transport energy from large-scale motions to the microscale, where it can be dissipated. In the radial quiet wind (blue trace), where the turbulence level is substantially lower, an enhancement of wave power near the ion cyclotron frequency is observed. In the active jet wind (black trace), a steep spectrum is seen at the plasma ion inertial and gyroscales, indicating a transition to kinetic range turbulence and possibly the dissipation of turbulent energy to

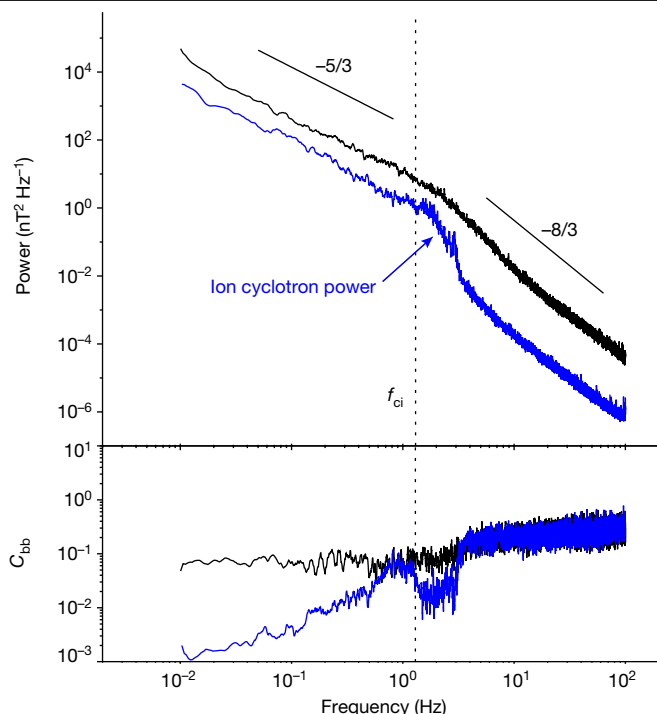


Fig. 4 | Power spectral density and magnetic compressibility of magnetic field fluctuations in quiet and jet wind. **a**, Thirty-minute integrated power spectra of fluctuations in quiet (blue) and jet (black) solar wind conditions show the transition from a magnetohydrodynamic inertial range to dissipation- or dispersion-range turbulence, here compared to the spacecraft-frame frequency $f^{-5/3}$ and $f^{-8/3}$ power laws. The quiet wind spectrum (blue) shows increased power near the ion cyclotron frequency (f_{ci} ; dashed vertical line) that is associated with enhanced magnetic helicity (Fig. 3e). **b**, The ratio of magnitude (B) to $\text{tr}(B)$ spectra indicates increased magnetic compressibility during jet intervals (black) compared to quiet wind (blue) up to the dissipation scale (a slope of $-8/3$). The ion cyclotron band corresponds to lower compressibility, as expected.

heat the solar wind as it expands to fill the heliosphere. In both types of wind, the power levels are several orders of magnitude larger than at 1 AU. The magnetic compressibility⁴¹—defined as $C_{bb} = (\delta B / \delta B)^2$ —shows an increase at high frequencies, as expected for kinetic range turbulence (Fig. 4b). At low frequencies, the compressibility is larger in jet than in quiet wind, but remains small (C_{bb} less than about 0.1), indicating that jet fluctuations have an enhanced compressible component but are still predominantly Alfvénic⁴⁴. In the quiet wind, the band of enhanced power near the cyclotron frequency has a reduced magnetic compressibility, as expected for quasi-parallel ion cyclotron waves⁴⁰.

PSP encounter 1 reveals a more structured and dynamic solar wind than is seen at 1 AU, with impulsive magnetic field reversals and plasma jets embedded in a quiet radial wind emerging from a small equatorial coronal hole. As PSP goes to lower altitudes, eventually to $9.8R_{\odot}$ during the upcoming solar maximum, we expect to descend below the altitude where the wind becomes super-Alfvénic ($v_R > v_A$) and measure the interface between the corona and the solar wind for the first time.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1818-7>.

- Parker, E. N. Dynamics of the interplanetary gas and magnetic fields. *Astrophys. J.* **128**, 664–676 (1958).
- Neugebauer, M. & Snyder, C. W. Solar plasma experiment. *Science* **138**, 1095–1097 (1962).
- McComas, D. J. et al. Weaker solar wind from the polar coronal holes and the whole Sun. *Geophys. Res. Lett.* **35**, L18103 (2008).
- Abbo, L. et al. Slow solar wind: observations and modeling. *Space Sci. Rev.* **201**, 55–108 (2016).
- Lapenta, G. & Knoll, D. A. Effect of a converging flow at the streamer cusp on the genesis of the slow solar wind. *Astrophys. J.* **624**, 1049–1056 (2005).
- Einaudi, G., Boncinelli, P., Dahlburg, R. B. & Karpen, J. T. Formation of the slow solar wind in a coronal streamer. *J. Geophys. Res.* **104**, 521–534 (1999).
- Fisk, L. A. & Schwadron, N. A. The behavior of the open magnetic field of the Sun. *Astrophys. J.* **560**, 425–438 (2001).
- Antiochos, S. K., Mikic, Z., Titov, V. S., Lionello, R. & Linker, J. A. A model for the sources of the slow solar wind. *Astrophys. J.* **731**, 112 (2011).
- Wang, Y.-M. & Sheeley, N. R. On potential field models of the solar corona. *Astrophys. J.* **392**, 310–319 (1992).
- Cranmer, S. R. Coronal holes. *Living Rev. Sol. Phys.* **6**, 3 (2009).
- Hollweg, J. V. & Johnson, W. Transition region, corona, and solar wind in coronal holes: some two-fluid models. *J. Geophys. Res.* **93**, 9547–9554 (1988).
- Verdini, A., Velli, M., Matthaeus, W. H., Oughton, S. & Dmitruk, P. A turbulence-driven model for heating and acceleration of the fast wind in coronal holes. *Astrophys. J. Lett.* **708**, 116–120 (2010).
- Parker, E. N. Heating solar coronal holes. *Astrophys. J.* **372**, 719–727 (1991).
- Cranmer, S. R., Field, G. B. & Kohl, J. L. Spectroscopic constraints on models of ion cyclotron resonance heating in the polar solar corona and high-speed solar wind. *Astrophys. J.* **518**, 937–947 (1999).
- Fox, N. J. et al. The Solar Probe Plus mission: humanity's first visit to our star. *Space Sci. Rev.* **204**, 7–48 (2016).
- Breneman, A. W. et al. STEREO and Wind observations of intense cyclotron harmonic waves at the Earth's bow shock and inside the magnetosheath. *J. Geophys. Res. Space Phys.* **118**, 7654–7664 (2013).
- Bale, S. D. et al. The FIELDS instrument suite for Solar Probe Plus. *Space Sci. Rev.* **204**, 49–82 (2016).
- Parker, E. N. Dynamical theory of the solar wind. *Space Sci. Rev.* **4**, 666–708 (1965).
- Altschuler, M. D. & Newkirk, G. Magnetic fields and structure of the solar corona: 1. Methods of calculating coronal fields. *Sol. Phys.* **9**, 131–149 (1969).
- Schatten, K. H., Wilcox, J. M. & Ness, N. F. A model of interplanetary and coronal magnetic fields. *Sol. Phys.* **6**, 442–455 (1969).
- Arge, C. N. & Pizzo, V. J. Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J. Geophys. Res. Space Phys.* **105**, 10465–10479 (2000).
- Riley, P. et al. A comparison between global solar magnetohydrodynamic and potential field source surface model results. *Astrophys. J.* **653**, 1510–1516 (2006).
- Hoeksema, J. T. *Structure and Evolution of the Large Scale Solar and Heliospheric Magnetic Fields*. PhD thesis, Stanford Univ. (1984).
- Lee, C. O. et al. Coronal field opens at lower height during the solar cycles 22 and 23 minimum periods: IMF comparison suggests the source surface should be lowered. *Sol. Phys.* **269**, 367–388 (2011).
- Riley, P. et al. Predicting the structure of the solar corona and inner heliosphere during Parker Solar Probe's first perihelion pass. *Astrophys. J. Lett.* **874**, 15 (2019).
- Levine, R. H., Altschuler, M. D., Harvey, J. W. & Jackson, B. V. Open magnetic structures on the Sun. *Astrophys. J.* **215**, 636–651 (1977).
- Kasper, J. C. et al. Solar Wind Electrons Alphas and Protons (SWEAP) investigation. *Space Sci. Rev.* **204**, 131 (2016).
- McComas, D. J. et al. Probing the energetic particle environment near the Sun. *Nature* <https://doi.org/10.1038/s41586-019-1811-1> (2019).
- Belcher, J. W. & Davis, L. Large-amplitude Alfvén waves in the interplanetary medium, 2. *J. Geophys. Res.* **76**, 3534–3563 (1971).
- Howes, G. G. et al. The slow mode nature of compressible wave power in solar wind turbulence. *Astrophys. J. Lett.* **753**, 19 (2012).
- Horbury, T. S. et al. Short, large-amplitude speed enhancements in the near-Sun fast solar wind. *Mon. Not. R. Astron. Soc.* **478**, 1980–1986 (2018).
- Gosling, J. T., Tian, H. & Phan, T. D. Pulsed Alfvén waves in the solar wind. *Astrophys. J. Lett.* **737**, 35 (2011).
- Balogh, A., Forsyth, R. J., Lucek, E. A., Horbury, T. S. & Smith, E. J. Heliospheric magnetic field polarity inversions at high heliographic latitudes. *Geophys. Res. Lett.* **26**, 631–634 (1999).
- Yamauchi, Y., Moore, R. L., Suess, S. T., Wang, H. & Sakurai, T. The magnetic structure of Ho macropicules in solar coronal holes. *Astrophys. J.* **605**, 511–520 (2004).
- Raouafi, N.-E. & Stenborg, G. Role of transients in the sustainability of solar coronal plumes. *Astrophys. J.* **787**, 118 (2014).
- Roberts, M. A., Uritsky, V. M., DeVore, C. R. & Karpen, J. T. Simulated encounters of the Parker Solar Probe with a coronal-hole jet. *Astrophys. J.* **866**, 14 (2018).
- Meyer-Vernet, N., Issautier, K. & Moncuquet, M. Quasi-thermal noise spectroscopy: the art and the practice. *J. Geophys. Res. Space Phys.* **122**, 7925–7945 (2017).
- Schekochihin, A. A. et al. Astrophysical gyrokinetics: kinetic and fluid turbulent cascades in magnetized weakly collisional plasmas. *Astrophys. J.* **182**, 310–377 (2009).
- Matthaeus, W. H., Goldstein, M. L. & Smith, C. Evaluation of magnetic helicity in homogeneous turbulence. *Phys. Rev. Lett.* **48**, 1256–1259 (1982).
- Jian, L. K. et al. Electromagnetic waves near the proton cyclotron frequency: STEREO observations. *Astrophys. J.* **786**, 123 (2014); erratum **847**, 1 (2017).
- Bruno, R. & Carbone, V. The solar wind as a turbulence laboratory. *Living Rev. Sol. Phys.* **10**, 2 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

¹Space Sciences Laboratory, University of California, Berkeley, CA, USA. ²Physics Department, University of California, Berkeley, CA, USA. ³The Blackett Laboratory, Imperial College London, London, UK. ⁴School of Physics and Astronomy, Queen Mary University of London, London, UK. ⁵Smithsonian Astrophysical Observatory, Cambridge, MA, USA. ⁶School of Physics and Astronomy, University of Minnesota, Minneapolis, MN, USA. ⁷Department of Physics & Astronomy, University of New Hampshire, Durham, NH, USA. ⁸Space Science Center, University of New Hampshire, Durham, NH, USA. ⁹Department of Physics, University of Maryland, College Park, MD, USA. ¹⁰Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA. ¹¹Joint Space Science Institute, University of Maryland, College Park, MD, USA. ¹²LPC2E, University of Orléans, CNRS, Orléans, France. ¹³Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, USA. ¹⁴Code 695, NASA

Goddard Space Flight Center, Greenbelt, MD, USA. ¹⁵Goddard Planetary Heliophysics Institute, University of Maryland Baltimore County, Baltimore, MD, USA. ¹⁶Code 672, NASA Goddard Space Flight Center, Greenbelt, MD, USA. ¹⁷Department of Physics and Astronomy, University of Iowa, Iowa City, IA, USA. ¹⁸Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. ¹⁹Heliophysics Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA. ²⁰University of Applied Sciences and Arts Northwestern Switzerland, Windisch, Switzerland. ²¹LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, Meudon, France. ²²Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. ²³Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA. ²⁴Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA, USA. *e-mail: bale@berkeley.edu

Methods

Heliocentric RTN coordinates

We use so-called heliocentric RTN coordinates in our study. R points from the Sun centre to the spacecraft. T lies in the spacecraft plane (close to the ecliptic) and is defined as the cross product of the solar rotation axis with R. T points in the direction of prograde rotation. N completes a right-handed system.

Statistical modes

To examine the large-scale magnetic structure (Fig. 1b), we seek to remove the rapidly varying spikes observed in Fig. 1a. To do this we produce statistical modes which are defined by binning the full cadence magnetic field observations into one-hour intervals and for each interval, calculating the modal value—the peak of the histogram of field values within each interval.

Identification of jet intervals

We calculate that approximately 6% of the duration of encounter 1 consists of jet intervals. That number is computed by measuring the duration of positive polarity B_r intervals (58,973 s) occurring from 30 October 2018 to 11 November 2018 UT (103,6800 s total). This interval was chosen to correspond to the negative polarity interval centred on D (Fig. 1b), which has primarily negative polarity over the coronal hole, and does not have transient coronal mass ejection events. The positive-polarity jets were identified using a simple zero-crossing algorithm applied to 1-s cadence radial magnetic field data B_r . Of course, not all so-called jets contain full polarity reversals. Biasing this calculation with an amplitude offset will produce a larger fraction of jet times; this is an ongoing study.

PSP/FIELDS measurement details

Measurements presented here were made by the FIELDS¹⁷ and SWEAP²⁷ instruments on the PSP spacecraft. Magnetic field measurements in Fig. 1a were made by the FIELDS fluxgate magnetometer and are averaged to 1-s cadence from their native cadence, which varies from about 2.3 to 293 samples per second over encounter 1. The B_r data shown in Fig. 1b are derived from the 1-s data by computing the distribution of amplitudes in one-hour intervals with an amplitude resolution of 1 nT, and by finding the peak value of that distribution: that is, the statistical mode. This technique removes the fluctuating ‘jet’ intervals, without introducing the amplitude bias of an averaging algorithm.

The magnetic field measurements in Fig. 2 start at 1-s cadence, averaged down from their native cadence as described above. All magnetic field measurements here are calibrated accurate to better than 0.5 nT. SWEAP velocity measurements are made by the Solar Probe Cup (SPC) sensor at a cadence of about 1 measurement per 0.87 s and then averaged to 5-s intervals. The 1-s cadence magnetic field data are then averaged onto these 5-s time intervals. This reduces fluctuation noise in the SPC data and provides velocity and magnetic field measurements at the same cadence. The plasma density measurements in Fig. 1a are made using the FIELDS Low Frequency Receiver (LFR)⁴², which measures the fluctuating electric field across the V1-V2 antenna pair¹⁷ and computes the spectral density (also shown in Fig. 3a). The spectral peak is identified and associated with the electron plasma frequency f_{pe} , as described previously³⁷. Hence, the frequency of the peak amplitude gives a reliable estimate of the total plasma density. The spectral resolution of the LFR instrument is $\Delta f/f \approx 4\%$. The plasma frequency f_{pe} is proportional to $\sqrt{n_e}$, where n_e is the electron (total) density; therefore the resulting uncertainty in the density measurement is $\Delta n/n \approx 2\Delta f/f \approx 8\%$. Electric field measurements used to compute the radial Poynting flux in Fig. 2b are measured directly as differential voltage pairs⁴³ between the V1-V2 and V3-V4 antennas¹⁷ and are then calibrated to electric field units by comparison to $-\mathbf{v} \times \mathbf{B}$ as computed from the SPC velocity and fluxgate magnetometer data. This enables us to

remove spacecraft offset electric fields and compute an effective probe-separation length, a standard technique used to calibrate electric field instrumentation⁴⁴. The electric field measurement is accurate to approximately 1 mV m⁻¹.

Measurements in Fig. 3a show the full spectrum of the RFS/LFR⁴² receiver in spectrogram form, as measured by the V1-V2 antenna pair. Wave intensity in Fig. 3a ranges from about 6×10^{-17} to 1.4×10^{-10} V² Hz⁻¹ and is represented logarithmically. The spectral bandwidth of the LFR receiver is $\Delta f/f = 4.5\%$ and the cadence of the measurement is 1 spectrum per 7 s. Figure 3b shows the electric field spectrogram of differential voltage measurements on the V1-V2 antenna pair from the Digital Fields Board (DFB) subsystem⁴³, with intensity in arbitrary logarithmic amplitude units. The spectral resolution of this channel of the DFB is $\Delta f/f \approx 6-12\%$ and the measurement cadence is 1 spectrum per 5.5 s. Figure 3c shows the magnetic field spectrogram of search coil magnetometer measurements from DFB⁴³, with intensity in arbitrary logarithmic amplitude units. The wavelet spectrogram in Fig. 3d and magnetic helicity spectrum in Fig. 3e were computed using the *wav_data* routine for the IDL programming language in the SPEDAS⁴⁵ suite of IDL analysis routines. Wave intensity in Fig. 3c is represented in logarithmic power in arbitrary units and is divided by a factor $P_K \propto f^{-5/3}$ (flattened), such that a power spectrum with spectral index $-5/3$ has no frequency dependence.

PFSS modelling and connection to PSP

Modelling the magnetic field time series (Fig. 1b) and tracing field lines from PSP down into the corona (Fig. 1c, d) was performed in two main steps.

(1) PFSS implementation. PFSS^{9,19,20} modelling used the recent open-source Python implementation *pfsspy*^{46,47}. This code package is freely available online, extremely flexible with regard to changing the input parameters and efficient (a full PFSS solution can be extracted in about 14 s including downloading the magnetogram on demand). Given a magnetogram and source-surface radius (R_{ss}) as boundary conditions, the code solves the Laplace equation (equation (1)) for the magnetic scalar potential, Φ_b , and outputs a full three-dimensional magnetic field within the annular volume bounded by the photosphere and the source-surface parameter. The choice of magnetogram data and values of source-surface height depicted in Fig. 1 are discussed in Methods section ‘Choice of magnetogram data and source-surface radius for Fig. 1’ below.

$$\nabla^2 \Phi_b(r) = 0 \quad (1)$$

(2) Ballistic propagation. The procedure to magnetically connect PSP to a particular location at the outer boundary of the PFSS solution domain follows refs. ^{48–50}, where the field line intersecting the position of PSP is assumed to follow a Parker spiral¹ with a curvature determined by the co-temporal solar wind velocity measurement at that position. As discussed in ref. ⁴⁸, although at lower radii this approximation is strongly perturbed by both corotational effects and the acceleration of the solar wind, these effects actually shift the coronal longitude by a similar magnitude but in opposite directions, resulting in an estimated error in longitude of less than 10°. This produces a very simple mapping (equation (2)) from spacecraft spherical Carrington coordinates (r_{PSP} , θ_{PSP} , ϕ_{PSP}) to coordinates on the source surface (r , θ , ϕ), involving the solar sidereal rotation rate, Ω_s , the measured solar wind speed, v_R , and the source-surface height R_{ss} .

$$\begin{pmatrix} r \\ \theta \\ \phi \end{pmatrix} = \begin{pmatrix} R_{ss} \\ \theta_{PSP} \\ \phi_{PSP} + \frac{\Omega_s}{v_R} (r_{PSP} - R_{ss}) \end{pmatrix} \quad (2)$$

To generate time series predictions, we first download a magnetogram, choose a source-surface height and generate a PFSS solution using equation (1). We then take the trajectory of PSP and use equation (2) to produce a time series of latitudes and longitudes on the source surface to which PSP was connected (see red and blue trajectories in Fig. 1c, d, Extended Data Figs. 1–3). For each latitude and longitude we obtain a B_R value at the source surface from the PFSS model. Finally, we scale each B_R value by $C(R_{SS}/r_{PSP})^2$ to produce an estimate of B_R at the location of PSP as a function of time. C is an empirically determined constant used to scale the time series prediction to match the peak measured magnetic field. Its value is dependent on the choice of magnetogram but approaches unity as the source-surface radius decreases and more flux is opened to the heliosphere. For the model results shown in Fig. 1, the values of C are 6.7 ($2.0R_\odot$ model) and 1.4 ($1.2R_\odot$ model).

To produce field line traces and generate Fig. 1c, d, we start with the time series of latitudes and longitudes on the source surface connected to PSP. For each pair of coordinates, we use *pfsspy*'s built-in field line tracer. Given the output of the *pfsspy* model, we supply the source-surface latitudes and longitudes and the field line tracer generates a field line which starts from that point and propagates it down to the photosphere. The model also provides a polarity for each field line generated, and we use these to colourize the field lines plotted in Fig. 1c, d.

Choice of magnetogram data and source-surface radius for Fig. 1

Synoptic maps of the photospheric magnetic field are available from multiple sources, which give variable outputs from the PFSS model. In this work we consider the NSO/GONG zero-point-corrected data product⁵¹, SDO/HMI vector magnetogram data product⁵², and the DeRosa (LMSAL) modelled magnetogram⁵³. GONG has the advantage of being operationally certified for space weather predictions, SDO/HMI is space-based and offers better resolution, and the DeRosa model assimilates HMI data and uses a surface flux transport and far-side helioseismological data to simulate photospheric dynamics such as differential rotation.

Additional variation in the outputs of the PFSS model arises from the time evolution of photospheric observations. Synoptic magnetograms are built by many observations of the Sun from Earth as it rotates with an approximate 27-day period. Typically, only $\pm 60^\circ$ longitude about the central meridian (sub-Earth point) are used for each observation (the grey regions in Extended Data Figs. 1–3). Although these maps can be updated with new data as frequently as observations are made, parts of the Sun facing away from Earth cannot be updated until they rotate into view, meaning all synoptic maps consist of a mix of old and new data and evolve in time.

Finally, the model output depends considerably on the choice of the source-surface radius parameter (R_{SS}). The inferred structure at the source surface changes as the source surface is lowered; implied structure such as the PIL—the contour of $B_R = 0$ —becomes more structured and warped. The foot points of open field lines at the photosphere encompass larger areas, increasing the predicted size of coronal holes, and also increasing the total amount (both positive and negative) of magnetic flux crossing the source surface.

Our approach to make robust conclusions is to generate model results for multiple times from all three magnetogram sources for varying source surface radii. Colour maps of B_R at the source surface and the associated PILs are shown in Extended Data Figs. 1–3. The majority of models at $2.0R_\odot$ and below predict polarity inversions in the vicinity of 240° and 310° longitude at all source-surface radii, with additional polarity inversions around 10° and 140° longitude that develop at lower source-surface radii. These features are all consistent with PSP measurements; we emphasize that they are largely independent of the time of observation and the choice of magnetogram source. Although the

established²³ value of $2.5R_\odot$ still gives good results from a GONG evaluation, both the HMI and DeRosa models produce strong disagreement around the time of perihelion. In Fig. 1b–d we show results from the GONG zero-point-corrected map evaluated on 6 November 2018 UT about which our time range of analysis is symmetric. This evaluation shows all the above features and produces good time series agreements. We show source-surface radii of $2.0R_\odot$ and $1.2R_\odot$. These lower source-surface radii do have modern precedent: $2.0R_\odot$ is consistent with previous PFSS modelling done for the same interval²⁵, where that radius was chosen to better match the observed extent of coronal holes. Another previous work²⁴ investigated the impact of lowering the source surface radius on model results, observing that at solar minimum a lower ($<2.0R_\odot$) source-surface radius was required to populate equatorial coronal holes with open field lines and improve their estimates of magnetic field strength at 1 AU.

Data availability

The data used in this study are available at the NASA Space Physics Data Facility (SPDF), <https://spdf.gsfc.nasa.gov/index.html>.

42. Pulupa, M. et al. The Solar Probe Plus Radio Frequency Spectrometer: measurement requirements, analog design, and digital signal processing. *J. Geophys. Res. Space Phys.* **122**, 2836–2854 (2017).
43. Malaspina, D. M. et al. The Digital Fields Board for the FIELDS instrument suite on the Solar Probe Plus mission: analog and digital signal processing. *J. Geophys. Res. Space Phys.* **121**, 5088–5096 (2016).
44. Mozer, F. S. DC and low-frequency double probe electric field measurements in space. *J. Geophys. Res. Space Phys.* **121**, 10942–10953 (2016).
45. Angelopoulos, V. et al. The Space Physics Environment Data Analysis System (SPEDAS). *Space Sci. Rev.* **215**, 9 (2019).
46. Stansby, D. Dstansby/pfsspy: pfsspy 0.1.2 <https://zenodo.org/record/3237053#.Xcqc-1f7Suk> (2019).
47. Yeates, A. Antyeates1983/pfss: first release of pfss code <https://zenodo.org/record/1472183#.XcqdU1f7Suk> (2018).
48. Nolte, J. T. & Roelof, E. C. Large-scale structure of the interplanetary medium. I: High coronal source longitude of the quiet-time solar wind. *Sol. Phys.* **33**, 241–257 (1973).
49. Neugebauer, M., et al. Spatial structure of the solar wind and comparisons with solar data and models. *J. Geophys. Res.* **103**, 14587–14599 (1998).
50. Stansby, D., Horbury, T. S., Wallace, S. & Arge, C. N. Predicting large-scale coronal structure for Parker Solar Probe using open source software. *RNAAS* **3**, 57 (2019).
51. Clark, R., Harvey, J., Hill, F. & Toner, C. GONG magnetogram zero-point-correction status. *Bull. Am. Astron. Soc.* **35**, 822 (2003).
52. Hoeksema, J. T. et al. The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: overview and performance. *Sol. Phys.* **289**, 3483–3530 (2014).
53. Schrijver, C. J. & DeRosa, M. L. Photospheric and heliospheric magnetic fields. *Sol. Phys.* **212**, 165 (2003).

Acknowledgements The FIELDS experiment on the Parker Solar Probe spacecraft was designed and developed under NASA contract NNN06AA01C. The FIELDS team acknowledges the contributions of the Parker Solar Probe mission operations and spacecraft engineering teams at the Johns Hopkins University Applied Physics Laboratory. S.D.B. acknowledges the support of the Leverhulme Trust Visiting Professorship programme. Contributions from S.T.B. were supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program grant 80NSSC18K1201. This work uses data obtained by the Global Oscillation Network Group (GONG) programme, managed by the National Solar Observatory, which is operated by AURA, Inc. under a cooperative agreement with the National Science Foundation. The data were acquired by instruments operated by the Big Bear Solar Observatory, High Altitude Observatory, Learmonth Solar Observatory, Udaipur Solar Observatory, Instituto de Astrofísica de Canarias and Cerro Tololo Interamerican Observatory. D.B. was supported by UK STFC grant ST/P000622/1. J.P.E. and T.S.H. were supported by UK STFC grant ST/S000364/1. D.S. was supported by UK STFC grant ST/N000692/1. C.H.K.C. is supported by STFC Ernest Rutherford Fellowship number ST/N003748/2. T.D.W. and V.V.K. are supported by CNES.

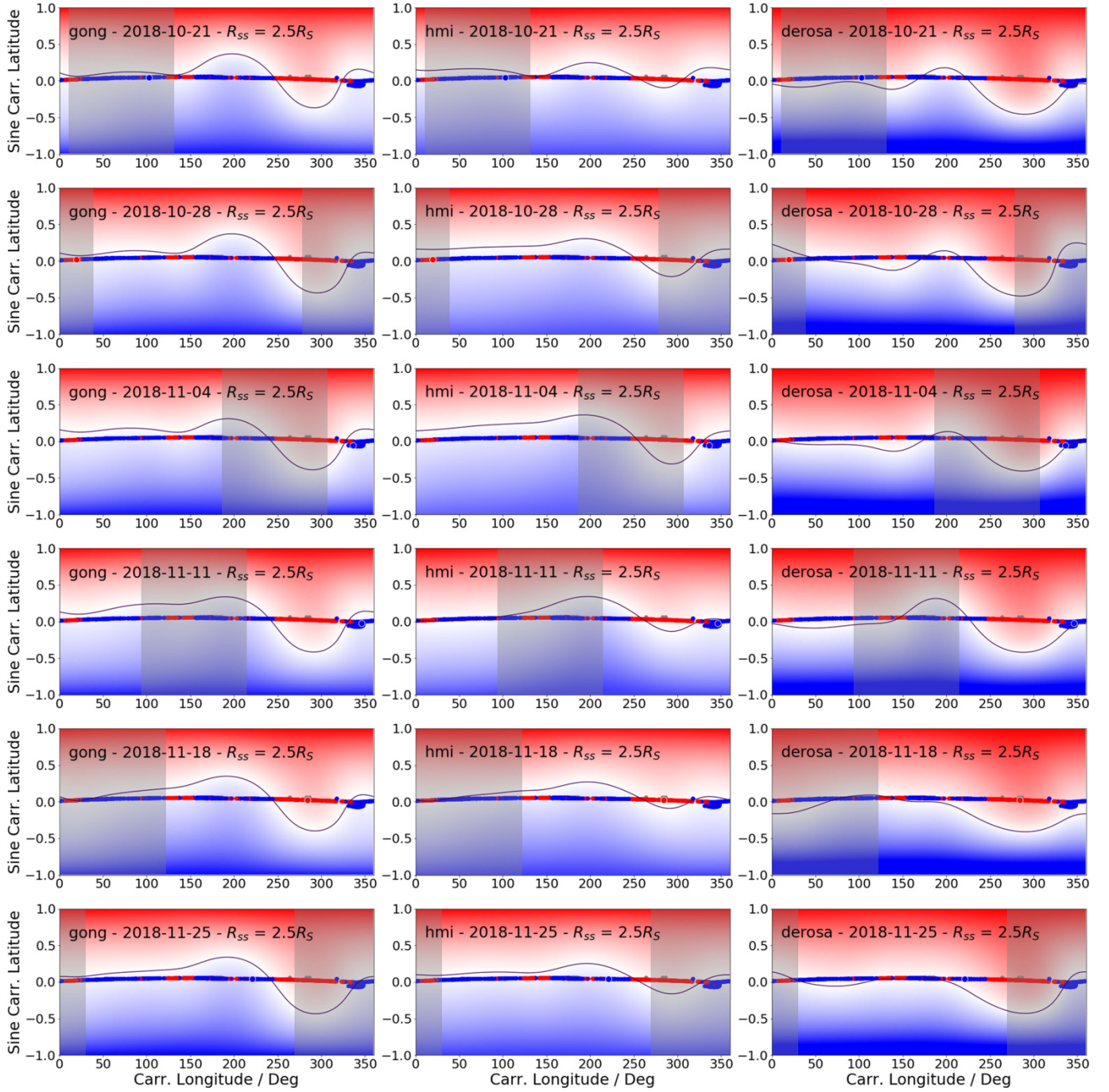
Author contributions S.D.B. wrote the manuscript with substantial contributions from S.T.B., B.D.G.C., C.H.K.C., T.S.H., M. Maksimovic, T.D.P. and M.V. All authors participated in the data interpretation, and read and commented upon the manuscript. S.D.B. led the FIELDS instrument team with contributions from J.W.B., T.A.B., T.D.W., K.G., P.R.H., D.E.L., R.J.M., M. Maksimovic, D.M.M., M.P. and N.E.R.

Competing interests The authors declare no competing interests.

Additional information

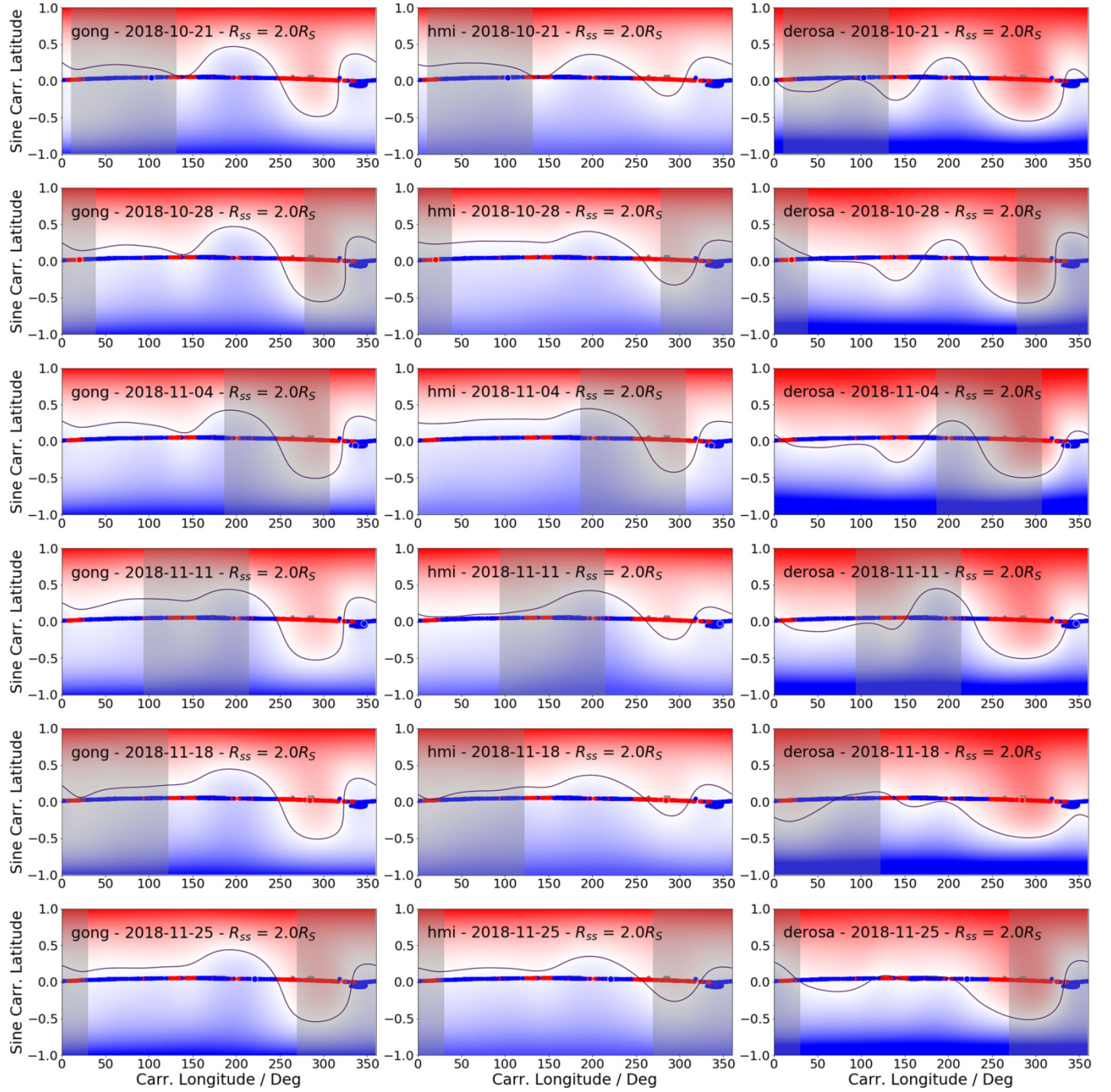
Correspondence and requests for materials should be addressed to S.D.B.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

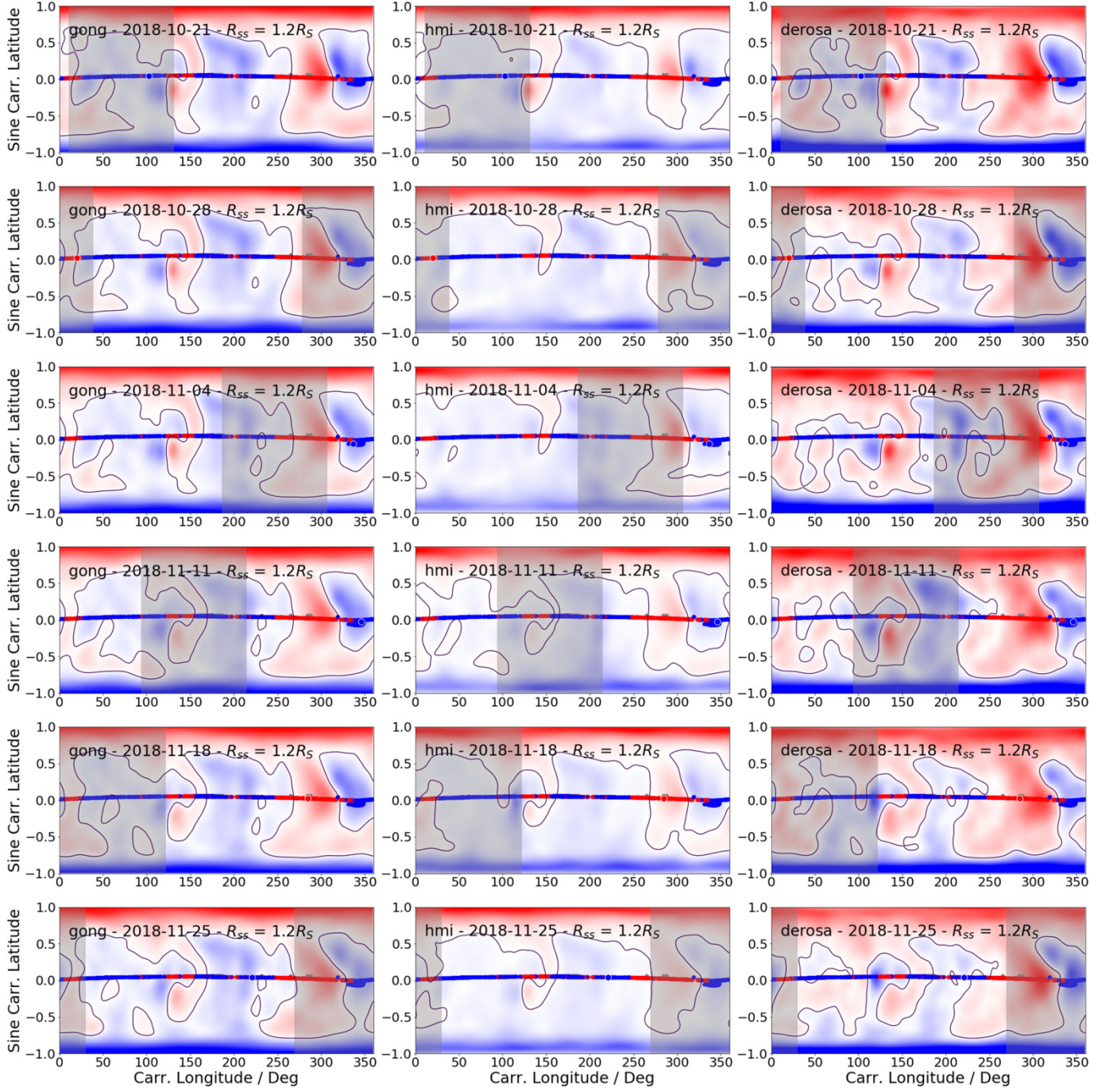


Extended Data Fig. 1 | Variation of PFSS neutral line topology with time and magnetogram choice at $R_{ss} = 2.5R_{\odot}$. Colour maps of B_r at the source surface from PFSS extractions with source-surface radius $R_{ss} = 2.5R_{\odot}$. Red indicates positive polarity and blue indicates negative polarity. The black line shows the PIL (the contour of $B_r = 0$). Superposed is the ballistically projected PSP trajectory coloured by the measured polarity. Perihelion occurred around 330° longitude. Left to right, the columns show extractions from the NSO/GONG,

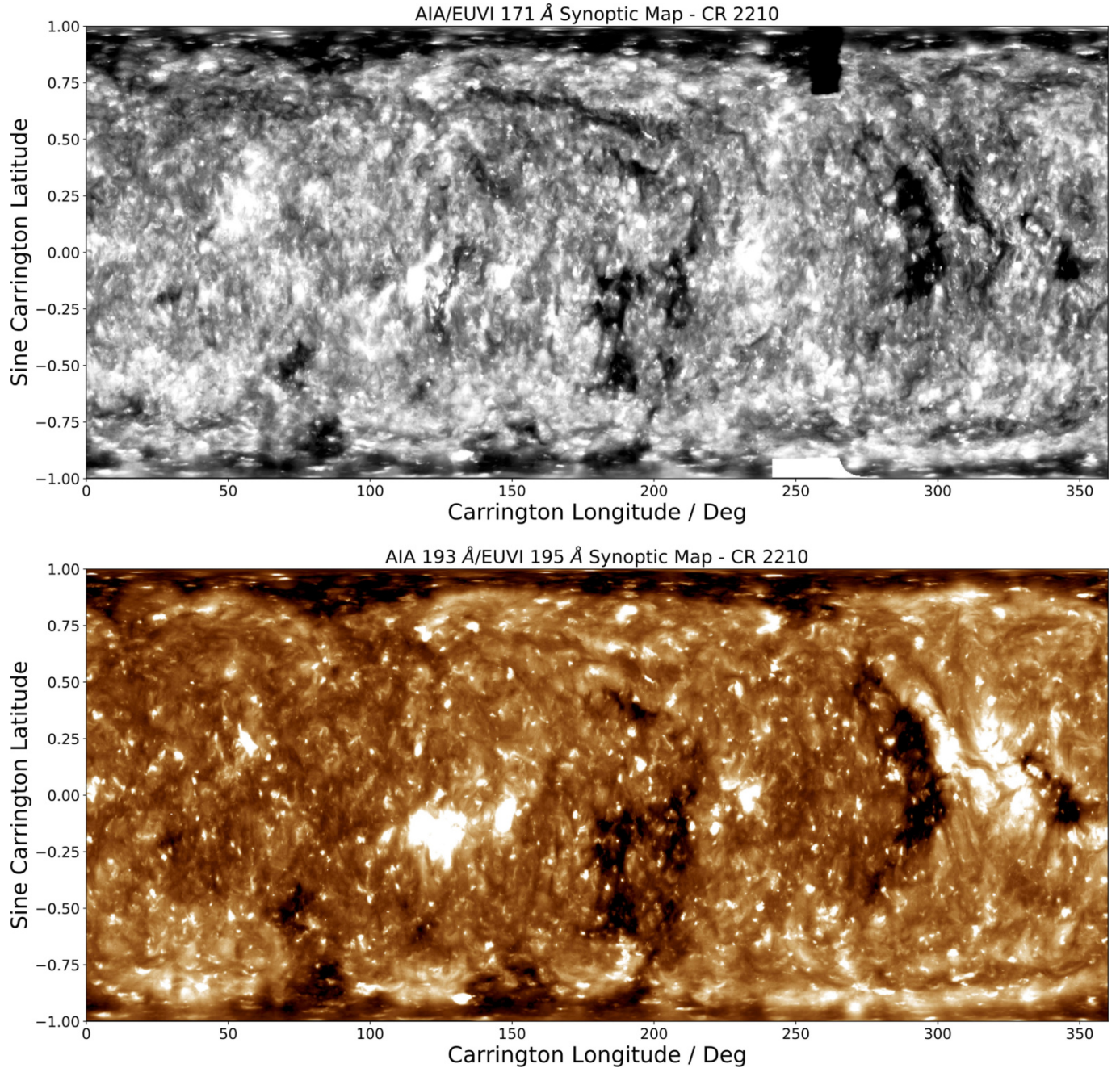
SDO/HMI and DeRosa LMSAL models. From top to bottom, the models are evaluated at a weekly cadence spanning six weeks about perihelion, with input magnetograms from each source taken as close in time as possible. The grey shading shows the region $\pm 60^\circ$ about the central meridian on the date of the model evaluation, indicating the portion of the Sun that could be observed at the time of observation.



Extended Data Fig. 2 | Variation of PFSS neutral line topology with time and magnetogram choice at $R_{ss} = 2.0R_{\odot}$. Colour maps of B_r at the source surface from PFSS extractions with $R_{ss} = 2.0R_{\odot}$. Other features are as described in Extended Data Fig. 1.

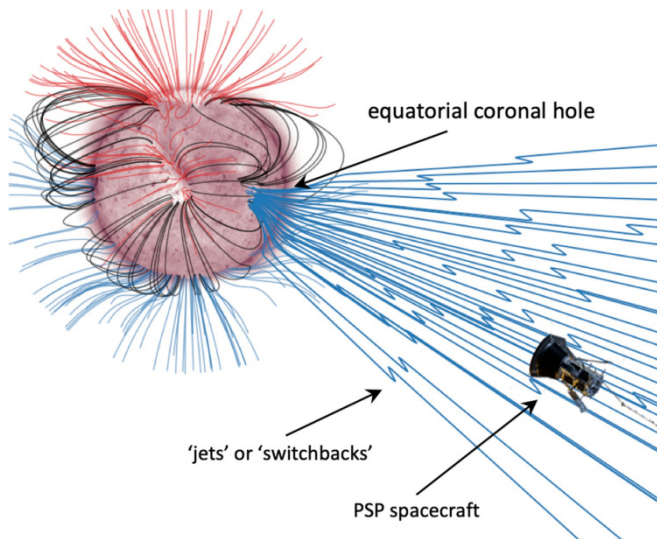


Extended Data Fig. 3 | Variation of PFSS Neutral Line topology with time and magnetogram choice at $R_{ss} = 1.2R_{\odot}$. Colour maps of B_r at the source surface from PFSS extractions with $R_{ss} = 1.2R_{\odot}$. Other features are as described in Extended Data Fig. 1.



Extended Data Fig. 4 | Synoptic maps of extreme-ultraviolet coronal emission from Carrington rotation 2,210, assembled from the STEREO-A/EUVI and SDO/AIA instruments. Top, 171-Å data showing coronal Fe IX emission at around 600,000 K. This is the background of Fig. 1c, d. Bottom, 193-Å (AIA) and 195-Å (EUVI) data showing emission from coronal Fe XII emission at around 1,000,000 K. The brightness is positively correlated with

the integrated plasma density squared along the line of sight. The dark regions in both images are probable locations of coronal holes, which are threaded by open magnetic field lines that allow plasma to evacuate into interplanetary space, resulting in under-dense regions. Carrington rotation 2,210 occurred from 20:51 26 October 2018 UT to 04:11 23 November 2018 UT.



Extended Data Fig. 5 | During encounter 1, PSP connected magnetically to a small negative-polarity equatorial coronal hole. This schematic shows a potential field extrapolation of the solar magnetic field at the time of the first perihelion pass of PSP. The solar surface is shown, coloured by AIA 211-Å extreme-ultraviolet emission (see Extended Data Fig. 4 for other wavelengths). Coronal holes appear as a lighter shade. Superposed are various field lines initialized at the solar disk. Black lines indicate closed loops, blue and red illustrate open field lines with negative and positive polarities, respectively. As depicted here and in Fig. 1c, d, at perihelion PSP connected to a negative equatorial coronal hole. The 'switchbacks' (the jets) observed by PSP (Fig. 1a) are illustrated as kinks in the open field lines emerging from the coronal hole that connect to PSP. (Note that neither the radial distance to the spacecraft nor the scale or amplitude of the jets or switchbacks are to scale.) Spacecraft graphic is courtesy of NASA/Johns Hopkins APL.

Phonon heat transfer across a vacuum through quantum fluctuations

<https://doi.org/10.1038/s41586-019-1800-4>

King Yan Fong^{1,3}, Hao-Kun Li^{1,3}, Rongkuo Zhao¹, Sui Yang¹, Yuan Wang¹ & Xiang Zhang^{1,2*}

Received: 15 February 2019

Accepted: 17 October 2019

Published online: 11 December 2019

Heat transfer in solids is typically conducted through either electrons or atomic vibrations known as phonons. In a vacuum, heat has long been thought to be transferred by radiation but not by phonons because of the lack of a medium¹. Recent theory, however, has predicted that quantum fluctuations of electromagnetic fields could induce phonon coupling across a vacuum and thereby facilitate heat transfer^{2–4}. Revealing this unique quantum effect experimentally would bring fundamental insights to quantum thermodynamics⁵ and practical implications to thermal management in nanometre-scale technologies⁶. Here we experimentally demonstrate heat transfer induced by quantum fluctuations between two objects separated by a vacuum gap. We use nanomechanical systems to realize strong phonon coupling through vacuum fluctuations, and observe the exchange of thermal energy between individual phonon modes. The experimental observation agrees well with our theoretical calculations and is unambiguously distinguished from other effects such as near-field radiation and electrostatic interaction. Our discovery of phonon transport through quantum fluctuations represents a previously unknown mechanism of heat transfer in addition to the conventional conduction, convection and radiation. It paves the way for the exploitation of quantum vacuum in energy transport at the nanoscale.

Quantum mechanics states that quantum fields are never at rest but fluctuate constantly, even at a temperature of absolute zero. These fluctuations lead to extraordinary physical consequences in many areas, ranging from atomic physics (for example, spontaneous emission and the Lamb shift⁷) to cosmology (for example, Hawking radiation⁸). In 1948, Casimir described a force that acts between neutral objects based on quantum fluctuations of electromagnetic fields⁹. This force is of both fundamental interest in quantum field theory and practical importance in nanoscale and microscale technology^{10,11}. Although the mechanical consequences of the Casimir effect have been extensively studied and precisely quantified^{12–17}, its role in thermodynamics is rarely explored. Recently, it has been predicted that the Casimir effect can induce phonon transport between nearby objects and thus transfer heat through a vacuum gap^{2–4}. However, this intriguing quantum phenomenon has not been observed owing to stringent experimental requirements for nanometre gaps. At such small distances, other effects such as charge–charge interactions^{18,19}, evanescent electric fields²⁰ and surface phonon polaritons²¹ may contribute and obscure experimental verification.

Here we experimentally demonstrate heat transfer between two objects driven by quantum vacuum fluctuations. Using nanomechanical systems to access individual phonon modes and resonantly enhance the thermal energy exchange, we boost the distance range at which the phenomenon becomes observable by over two orders of magnitude to hundreds of nanometres, compared to the nanometre to sub-nanometre range predicted for bulk solids^{2–4}. This allows us to single out the Casimir effect from other short-range effects. We quantify

the temperature change of the phonon modes through their thermal Brownian motion and unambiguously show that the two phonon modes thermalize in the strong Casimir phonon coupling regime. Our result reveals a new mechanism of heat transfer through a quantum vacuum. It also opens up new opportunities for studying quantum thermodynamics and energy transport using nanomechanical devices.

To illustrate the concept, we consider the interaction of two phonon modes based on a spring-mass model (shown in Fig. 1a). Two objects attached to springs are linked to thermal baths at different temperatures and undergo thermal Brownian motions. Displacement of the two objects perturbs the zero-point energy of the electromagnetic vacuum, giving rise to the Casimir interaction⁹. In the regime in which thermal Brownian motions of the objects are much slower than the response time of the Casimir interaction, the Casimir force acts instantaneously and is conservative in nature^{22–24}. The Casimir interaction effectively acts as a coupling spring that connects the two objects, through which the hot object agitates the cold object. As a result, thermal energy is transferred across the phonon modes from the hot to the cold side.

In the experimental setting, we use frequency-matched nanomechanical oscillators to realize and resonantly enhance this Casimir heat transfer effect (Fig. 1b). Two parallel membrane resonators, each clamped to a substrate at different temperatures (T_1 and T_2), are separated by an adjustable distance, d . In the presence of the Casimir force, $F_{\text{Cas}}(d)$, the system can be modelled as two coupled harmonic oscillators driven by Langevin forces from different temperature baths^{25,26}: $\ddot{u}_i + 2\gamma_i\dot{u}_i + \Omega^2 u_i - 2\Omega g_c(u_i - \alpha_i u_j) = \delta F_i/m_i$, where $i, j \in \{1, 2\}$, $i \neq j$,

¹Nanoscale Science and Engineering Center, University of California, Berkeley, CA, USA. ²Faculties of Science and Engineering, The University of Hong Kong, Hong Kong, China. ³These authors contributed equally: King Yan Fong, Hao-Kun Li. *e-mail: xiang@berkeley.edu

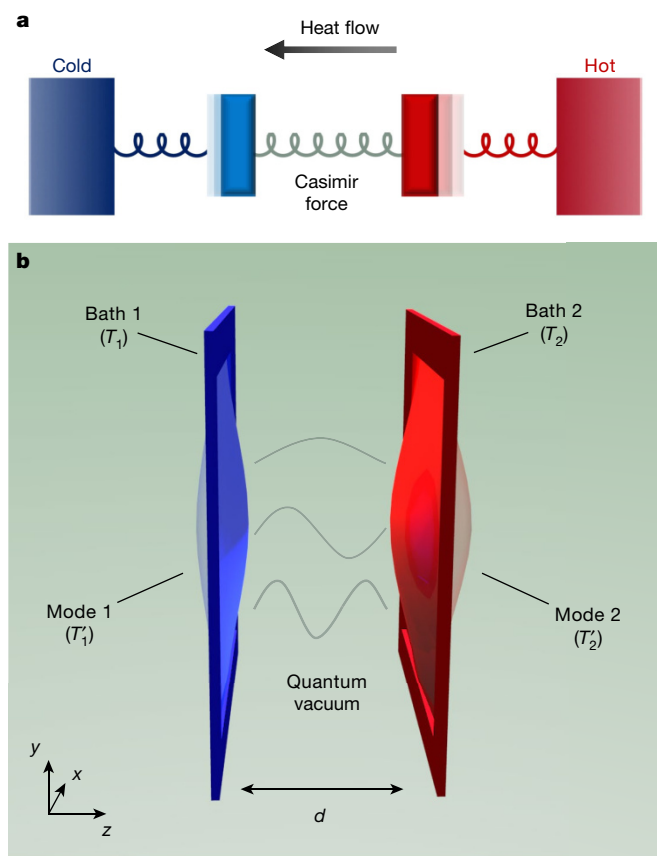


Fig. 1 | Casimir heat transfer driven by quantum vacuum fluctuations. **a**, As a conceptual illustration, we consider a spring-mass model in which two objects are separately linked to a hot and a cold thermal bath. The hot (or cold) object has higher (or lower) thermal energy and therefore undergoes greater (or lesser) thermal Brownian motion. Owing to the Casimir interaction, the two objects are effectively linked by a coupling spring through which the rapid thermal motion of the hot object agitates the cold object. As a result, thermal energy is transferred from the hot to the cold side. **b**, In the experimental setting, we use a pair of nanomechanical membrane resonators to demonstrate this mechanism of heat transfer. The two phonon modes (the fundamental modes of the membranes) have mode temperatures (T'_i) that are determined by their thermal Brownian motions. The Casimir interaction facilitates thermal energy exchange between the two phonon modes at short distances, d . As a result, the mode temperatures deviate from their bath temperatures ($T'_i \neq T_i$).

$g_c = F'_{\text{Cas}}(d)/2\Omega\rho_A$ is the coupling rate that arises from the Casimir force, and u_i , Ω , m_i , γ_i , α_i , ρ_A and δF_i are respectively the displacement, resonance frequency, effective mass, dissipation rate, mode-matching factor, membrane area density and Langevin force. At large separation, where the Casimir interaction is negligible, the phonon modes of the membranes are in thermal equilibrium with their respective thermal baths, that is, $T'_i = T_i$, where $k_B T'_i = m_i \Omega^2 \langle u_i^2 \rangle$ is the mode temperature determined by the thermal Brownian motion²⁷. At short distances, the Casimir interaction dominates and induces thermal energy exchange between the phonon modes, manifested as an observable deviation of the mode temperatures from their bath temperature (see Supplementary Information section 1).

We use optical interferometry to measure the thermal Brownian motion in order to determine the phonon mode temperatures (Fig. 2a). Using minimal laser power (8 μW) to avoid thermo-optical heating, we resolve the thermomechanical noise of the fundamental modes with a signal-to-background ratio of about 20 dB (Fig. 2e, f). The two high-stress stoichiometric Si_3N_4 membranes of different dimensions ($330 \times 330 \times 0.1 \mu\text{m}^3$ and $280 \times 280 \times 0.1 \mu\text{m}^3$) are coated with gold (75 nm) on both sides for the purposes of optical reflection and

electrical contact (Fig. 2b, c). The dimensions of the two membranes are different such that their fundamental flexural mode frequencies can be matched at different temperatures by thermally tuning the membrane stress. At bath temperatures $T_1 = 287.0 \text{ K}$ and $T_2 = 312.5 \text{ K}$, the resonances match at $\Omega/2\pi = 191.6 \text{ kHz}$ (Fig. 2d), with high quality factors of $Q_1 = 4.5 \times 10^4$ and $Q_2 = 2.0 \times 10^4$. A bias voltage (V_b) is applied across the two membranes to compensate for any built-in electrostatic potential that may overwhelm the Casimir effect.

An essential experimental requirement here is to align the two planar resonators with a high degree of parallelism, which has been a hurdle for precision measurement of the Casimir force between planar structures¹³. To solve the problem, we implement high-precision (below 10^{-4} rad) membrane alignment using an optical interferometric technique and an electrical method²⁸ with specific mesa structures and electrode patterning (Fig. 2b, c) (see Methods). This allows us to explore the Casimir interaction between two parallel planes separated by an unprecedented distance of around 300 nm (ref.¹³).

We observe the Casimir heat transfer between the phonon modes of the membranes (Fig. 3a). The mode temperatures show a strong dependence on the distance. At large separations, the mode temperatures are the same as their thermal bath temperatures, while at small separations (less than 600 nm) they begin to deviate. As the distance is decreased further to below 400 nm, T'_1 and T'_2 become nearly identical, showing thermalization of the two phonon modes. Such a heat-transfer effect is observed only when the resonance frequencies are matched within the linewidth, that is, when $|\Omega_2 - \Omega_1|$ is less than γ_1, γ_2 . In the measurement, the mode temperatures are determined by their thermal Brownian motions. The mechanical motion can be decomposed into $u_i(t) = X_i(t)\cos\Omega t + Y_i(t)\sin\Omega t$, with $X_i(t)$ and $Y_i(t)$ being the two quadrature components. The two measured quadrature components display a circularly symmetric distribution in the phase space, showing that the thermal motions are random with all phases being equally available (Fig. 3b, c). A plot of the probability distribution of the total energy $E_i = m_i \Omega^2 (X_i^2 + Y_i^2)/2$ (Fig. 3d) shows that it follows the statistics of a canonical ensemble, that is, $P(E_i) \propto e^{-E_i/k_B T'_i}$. The difference between T_i and T'_i determines the net heat flux flowing from the thermal bath to the phonon mode, given by $P_i = 2\gamma_i k_B (T_i - T'_i)$. From the measured mode temperatures (Fig. 3a), we obtain the averaged heat flux transferred across the two thermal baths by $P_{2 \rightarrow 1} = (P_2 - P_1)/2$ (Fig. 3e).

The observed phenomenon can be quantitatively explained by the competition between the Casimir coupling rate (g_c) and the mode-bath thermal exchange rate ($\gamma_i = \Omega/2Q_i$). When d decreases from 600 nm to 350 nm, g_c increases rapidly and the system evolves from weak ($g_c \ll \gamma_i$) to strong ($g_c \gg \gamma_i$) Casimir phonon coupling regime (Extended Data Fig. 1). Using coupled-mode Langevin equations (see Supplementary Information, section 1), we derive the mode temperatures and the heat flux across the two thermal baths as:

$$T'_i = T_i + \frac{\gamma_j(T_j - T_i)}{(\gamma_i + \gamma_j) \left(1 + \frac{\gamma_i \gamma_j}{g_c^2}\right)}$$

$$P_{j \rightarrow i} = \frac{2\gamma_i \gamma_j k_B (T_j - T_i)}{(\gamma_i + \gamma_j) \left(1 + \frac{\gamma_i \gamma_j}{g_c^2}\right)}$$

where $g'_c = g_c \sqrt{\alpha_1 \alpha_2}$ is the effective coupling rate that accounts for mode mismatch. In the experiment, $\sqrt{\alpha_1 \alpha_2} = 0.97$ and therefore g'_c is approximately g_c . The theoretical prediction well describes the experimental data (solid lines in Fig. 3a, e). When $g_c \gg \gamma_i$, thermalization occurs ($T'_1 = T'_2$) and the heat flux reaches a maximum value of $6.5 \times 10^{-21} \text{ J s}^{-1}$. Additional experimental results obtained from different samples and conditions are presented in Extended Data Fig. 2. In our theoretical analysis, we apply the proximity force approximation²⁹, which is valid under the condition that the wavelength of the phonon mode is much

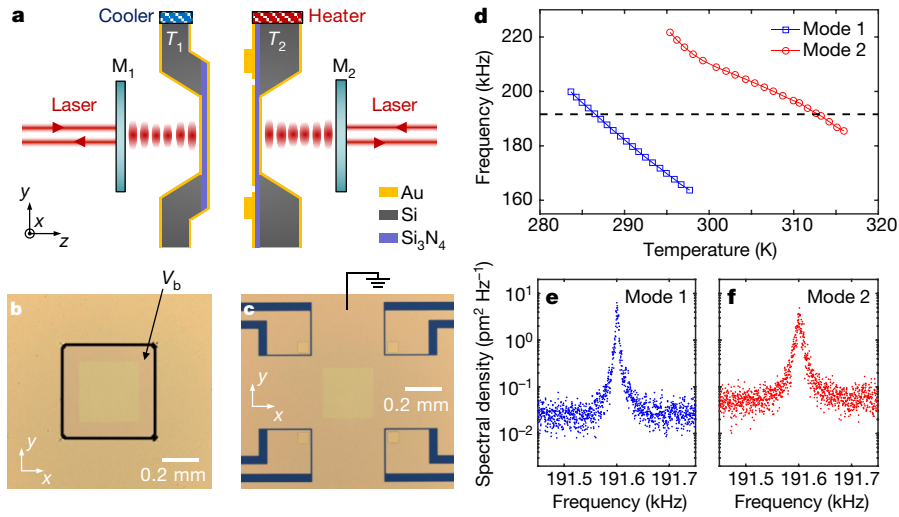


Fig. 2 | Experimental setup and fabricated samples. **a**, Cross-sectional view of the experimental setup, showing two nanomechanical Si_3N_4 membranes aligned in parallel and brought close together. Partially reflecting mirrors (M_1 and M_2) are placed behind the membranes, and laser beams (wavelength 633 nm) are sent from both sides to interferometrically measure the thermomechanical motion of the membranes. The distances between the mirrors and the membranes are feedback controlled by piezo-actuators to maintain long-term stability of the interferometric detection sensitivity. The two samples are mounted on a closed-loop thermoelectric cooler and heater to stabilize the sample temperatures and tune the mechanical resonance frequencies. The setup is kept in a vacuum below 10^{-6} Torr. **b, c**, Optical images

of the samples. The Si_3N_4 membranes (100 nm) are coated with gold (75 nm) on both sides for optical reflection and electrical contact. The mesa structure on the left sample (**b**) and the electrodes on the right sample (**c**) are fabricated for parallel alignment (see Methods). To compensate for the built-in electrostatic potential, a bias voltage V_b is applied between the membrane surfaces. **d**, At room temperature, resonance frequencies of the two modes differ by around 50 kHz owing to the difference in the dimensions of the membranes. By adjusting the sample temperature to tune the film stress, the two frequencies are matched at $\Omega/2\pi = 191.6$ kHz when bath temperatures reach $T_1 = 287.0$ K and $T_2 = 312.5$ K (black dashed line). **e, f**, Thermomechanical noise spectra of the fundamental modes.

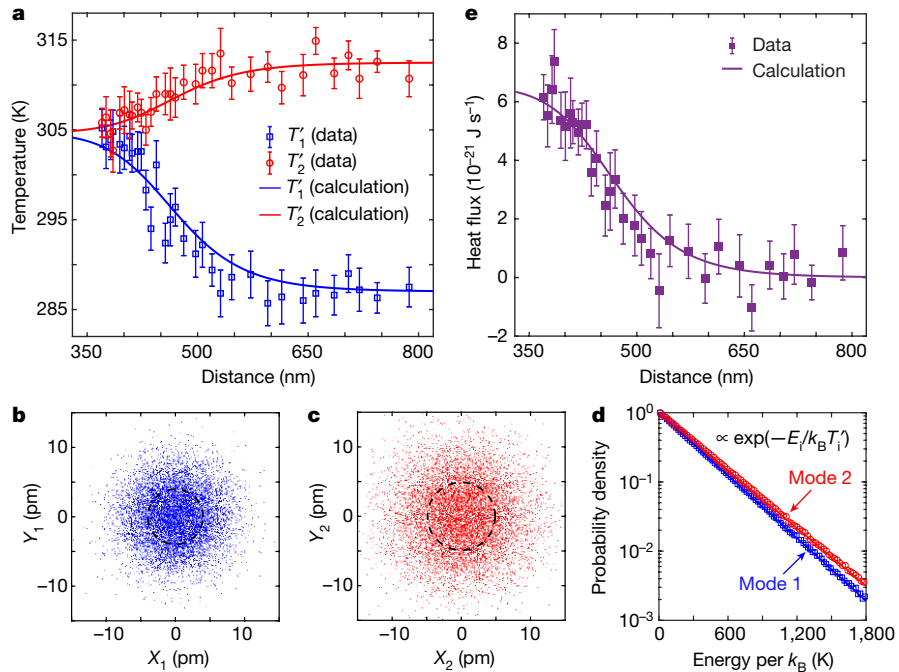


Fig. 3 | Observation of Casimir heat transfer. **a**, In the presence of the Casimir interaction, the mode temperatures deviate from their bath temperatures when the two membranes are brought close. At distances below 400 nm, T'_1 and T'_2 become nearly identical, showing thermalization of the two phonon modes. The mode temperatures are measured from the thermal Brownian motion, $k_B T'_i = m_i \Omega^2 \langle u_i^2 \rangle$. Error bars represent the standard error obtained from four hours of continuous measurement. The data agree well with calculations using coupled-mode Langevin equations (solid lines). **b, c**, Measured quadrature components (in picometres) of the thermal displacement of phonon modes 1 (**b**) and 2 (**c**) at $T'_1 = 287.0$ K and $T'_2 = 312.5$ K, respectively. Dashed lines indicate

standard deviations of the distributions; the enclosed areas are proportional to the mode temperatures. **d**, Probability distributions of the phonon-mode energy follow the statistics of a canonical ensemble $P(E_i) \propto e^{-E_i/k_B T'_i}$, represented by solid lines. **e**, Heat flux transferred across the two thermal baths as a function of distance, extracted from the measured mode temperatures in panel **a**. The error bars originate from error propagation in the calculation. The solid line represents the theoretical prediction of the coupled-mode model. Additional experimental results obtained from different samples and conditions are presented in Extended Data Fig. 2.

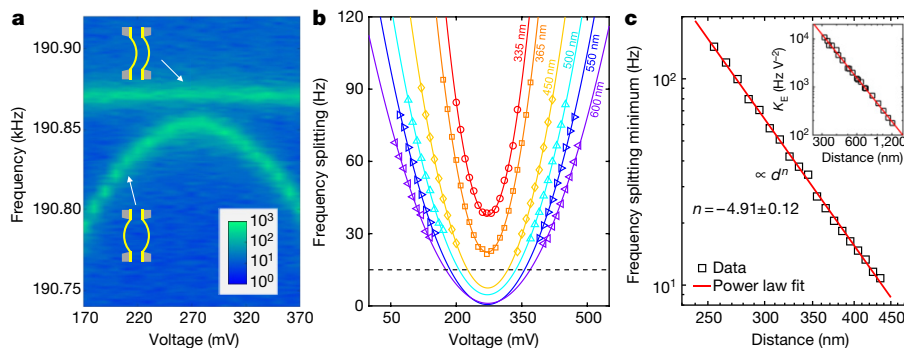


Fig. 4 | Strong phonon coupling through Casimir interaction.

a, Thermomechanical noise spectrum of membrane 2 at $d = 400$ nm. The two distinct branches result from strong coupling of the phonon modes. The upper (or lower) branch represents the even (or odd) mode where the two membranes move in the same (or opposite) direction. **b**, Frequency splitting of the thermomechanical noise spectrum shows a parabolic dependence on the bias voltage between the membranes (solid curves are parabolic fits). The dashed line represents the sum of the phonon mode linewidths, above which the frequency splitting can be well resolved in the spectra. The curvatures of the parabolas are proportional to the electrostatic interaction strength $K_E(d)$,

larger than the membrane separation (d/γ is of the order of 10^{-3} in the experiment). For higher-order phonon modes with wavelengths comparable to or smaller than the gap, modification of Casimir energy owing to the surface modulation of phonons needs to be considered^{2,30}.

To verify that the observed heat transfer is due to the Casimir interaction, we examine the phonon-mode coupling characteristics. When the membranes are brought close, we observe an anti-crossing feature in the thermal noise spectra, revealing strong coupling of the two modes (Fig. 4a). The frequency splitting Δf of the two peaks is a direct indicator of the coupling strength. Taking both the Casimir and the electrostatic effects into consideration, we can express the frequency splitting as $\Delta f = \Delta f_{\text{Cas}} + \Delta f_{\text{ele}}$, with $\Delta f_{\text{Cas}} = g'_c(d)/\pi$ and $\Delta f_{\text{ele}} = K_E(d)[(V_b - V_0)^2 + V_{\text{rms}}^2]$ being the Casimir and electrostatic components, respectively³¹. The coefficient $K_E(d)$ is proportional to d^{-3} , and V_0 and V_{rms} represent the first and second moments of the surface potential difference between the two metallized membranes. The unique dependence of Δf_{ele} on V_b and d (Fig. 4b) allows calibration of the absolute distance between the two membranes^{13,14,16} (see Methods). We observe that the surface potential V_0 remains constant as distance is varied (see Extended Data Fig. 3), which agrees with the theoretical prediction for parallel plane configuration³¹. When V_0 is compensated by the applied bias voltage V_b , we observe that the frequency splitting shows a distance dependence of $d^{-4.91 \pm 0.12}$ (Fig. 4c). This verifies that the Casimir effect dominates over the electrostatic effect in our measurement. This result also represents the first demonstration of strong phonon coupling induced by the Casimir force.

Finally, we distinguish the observed Casimir heat transfer from thermal radiation effects. Near-field thermal radiation generates heat flow through the vacuum gap and slightly modifies the temperature of the bulk membranes. This leads to changes in membrane stress and thus frequency shifts of the phonon modes (see Methods and Extended Data Fig. 4). Unlike Casimir phonon coupling, the thermal radiation effect does not depend on the frequency matching of the two modes, which we verify by offsetting the frequencies of the two modes through thermal tuning. The observed frequency shifts are less than 40 Hz in the distance range of our measurement. These frequency shifts correspond to temperature changes of less than 0.02 K, based on the measured frequency–temperature dependence of 2 kHz K⁻¹ for the membrane modes (Fig. 2d). The slight temperature changes agree with our calculation using the measured radiation heat transfer coefficient between gold surfaces³². On the other hand, thermal radiation pressure may also

provide mechanical coupling between two phonon modes. However, such an effect is estimated to be negligible in our experimental condition (less than 4% for distances shorter than 800 nm)³³ (see Extended Data Fig. 1e).

In conclusion, we have experimentally demonstrated heat transfer driven by quantum fluctuations using nanomechanical devices. Our observation is unambiguously distinguished from other effects, including electrostatic interactions and near-field thermal radiation. In this work, we have focused on heat transfer through single-phonon modes. When the majority of phonon modes in a solid take part in the thermal exchange process, the effect generalizes to heat transfer between two bulk solids^{2–4}. The ability to control thermal flow with a quantum vacuum opens up a new arena for studying quantum thermodynamics^{34,35} and for implementing quantum thermal machines⁵. Moreover, our method for achieving and controlling strong Casimir phonon coupling provides a versatile platform for implementing coherent phonon processes (for example, phonon state transfer and entanglement) using a quantum vacuum.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1800-4>.

- Bergman, T. L., Incropera, F. P., DeWitt, D. P. & Lavine, A. S. *Fundamentals of Heat and Mass Transfer* (John Wiley & Sons, 2011).
- Pendry, J. B., Sasiithlu, K. & Craster, R. V. Phonon-assisted heat transfer between vacuum-separated surfaces. *Phys. Rev. B* **94**, 075414 (2016).
- Ezzahri, Y. & Joulain, K. Vacuum-induced phonon transfer between two solid dielectric materials: illustrating the case of Casimir force coupling. *Phys. Rev. B* **90**, 115433 (2014).
- Budaev, B. V. & Bogoy, D. B. On the role of acoustic waves (phonons) in equilibrium heat transfer exchange across a vacuum gap. *Appl. Phys. Lett.* **99**, 053109 (2011).
- Terças, H., Ribeiro, S., Pezzutto, M. & Omar, Y. Quantum thermal machines driven by vacuum forces. *Phys. Rev. E* **95**, 022135 (2017).
- Chiloyan, V., Garg, J., Esfarjani, K. & Chen, G. Transition from near-field thermal radiation to phonon heat conduction at sub-nanometre gaps. *Nat. Commun.* **6**, 6755 (2015).
- Scully, M. O. & Zubairy, M. S. *Quantum Optics* Ch. 1 (Cambridge Univ. Press, 1997).
- Hawking, S. W. Particle creation by black holes. *Commun. Math. Phys.* **43**, 199–220 (1975).
- Casimir, H. B. G. On the attraction between two perfectly conducting plates. *Proc. K. Ned. Akad. Wet.* **51**, 793–795 (1948).
- Klimchitskaya, G. L., Mohideen, U. & Mostepanenko, V. M. The Casimir force between real materials: experiment and theory. *Rev. Mod. Phys.* **81**, 1827–1885 (2009).

11. Rodriguez, A. W., Capasso, F. & Johnson, S. G. The Casimir effect in microstructured geometries. *Nat. Photon.* **5**, 211–221 (2011).
12. Lamoreaux, S. K. Demonstration of the Casimir force in the 0.6 to 6 μm range. *Phys. Rev. Lett.* **78**, 5–8 (1997).
13. Bressi, G., Carugno, G., Onofrio, R. & Ruoso, G. Measurement of the Casimir force between parallel metallic surfaces. *Phys. Rev. Lett.* **88**, 041804 (2002).
14. Chan, H. B., Aksyuk, V. A., Kleiman, R. N., Bishop, D. J. & Capasso, F. Quantum mechanical actuation of microelectromechanical systems by the Casimir force. *Science* **291**, 1941–1944 (2001).
15. Munday, J. N., Capasso, F. & Parsegian, V. A. Measured long-range repulsive Casimir–Lifshitz forces. *Nature* **457**, 170–173 (2009).
16. Garcia-Sanchez, D., Fong, K. Y., Bhaskaran, H., Lamoreaux, S. & Tang, H. X. Casimir force and in situ surface potential measurements on nanomembranes. *Phys. Rev. Lett.* **109**, 027202 (2012).
17. Somers, D. A., Garrett, J. L., Palm, K. J. & Munday, J. N. Measurement of the Casimir torque. *Nature* **564**, 386–389 (2018).
18. Altfelder, I., Voevodin, A. A. & Roy, A. K. Vacuum phonon tunneling. *Phys. Rev. Lett.* **105**, 166101 (2010).
19. Xiong, S. et al. Classical to quantum transition of heat transfer between two silica clusters. *Phys. Rev. Lett.* **112**, 114301 (2014).
20. Prunnila, M. & Meltaus, J. Acoustic phonon tunneling and heat transport due to evanescent electric fields. *Phys. Rev. Lett.* **105**, 125501 (2010).
21. Shen, S., Narayanaswamy, A. & Chen, G. Surface phonon polaritons mediated energy transfer between nanoscale gaps. *Nano Lett.* **9**, 2909–2913 (2009).
22. Ford, L. H. & Vilenkin, A. Quantum radiation by moving mirrors. *Phys. Rev. D* **25**, 2569–2575 (1982).
23. Wilson, C. M. et al. Observation of the dynamical Casimir effect in a superconducting circuit. *Nature* **479**, 376–379 (2011).
24. Di Stefano, O. et al. Interaction of mechanical oscillators mediated by the exchange of virtual photon pairs. *Phys. Rev. Lett.* **122**, 030402 (2019).
25. Biehs, S.-A. & Agarwal, G. S. Dynamical quantum theory of heat transfer between plasmonic nanosystems. *J. Opt. Soc. Am. B* **30**, 700–707 (2013).
26. Barton, G. Classical van der Waals heat flow between oscillators and between half-spaces. *J. Phys. Condens. Matter* **27**, 214005 (2015).
27. Aspelmeyer, M., Kippenberg, T. J. & Marquardt, F. Cavity optomechanics. *Rev. Mod. Phys.* **86**, 1391–1452 (2014).
28. Ganjeh, Y. et al. A platform to parallelize planar surfaces and control their spatial separation with nanometer resolution. *Rev. Sci. Instrum.* **83**, 105101 (2012).
29. Blocki, J., Randrup, J., Świątecki, W. J. & Tsang, C. F. Proximity forces. *Ann. Phys.* **105**, 427–462 (1977).
30. Emig, T., Hanke, A., Golestanian, R. & Kardar, M. Normal and lateral Casimir forces between deformed plates. *Phys. Rev. A* **67**, 022114 (2003).
31. Kim, W. J., Sushkov, A. O., Dalvit, D. A. & Lamoreaux, S. K. Surface contact potential patches and Casimir force measurements. *Phys. Rev. A* **81**, 022505 (2010).
32. Shen, S., Mavrokefalos, A., Sambegoro, P. & Chen, G. Nanoscale thermal radiation between two gold surfaces. *Appl. Phys. Lett.* **100**, 233114 (2012).
33. Sushkov, A. O., Kim, W. J., Dalvit, D. A. R. & Lamoreaux, S. K. Observation of the thermal Casimir force. *Nat. Phys.* **7**, 230–233 (2011).
34. Kosloff, R. Quantum thermodynamics: a dynamical viewpoint. *Entropy* **15**, 2100–2128 (2013).
35. Vinjanampathy, S. & Anders, J. Quantum thermodynamics. *Contemp. Phys.* **57**, 545–579 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Numerical calculation of the Casimir force

We calculate the Casimir force between the membrane samples on the basis of Lifshitz theory³⁶. The calculation takes into account the finite conductivity and dispersion of the gold film and the geometry of the fabricated membrane structure (Extended Data Fig. 1a). The Casimir force per area can be written as $F_{\text{Cas}}(d) = -\eta \hbar c \pi^2 / 240 d^4$, where η is the correction factor applied to the Casimir force between two planar perfect conductors (Extended Data Fig. 1b). When the system enters the strong coupling regime ($g_c \gg \gamma_1, \gamma_2$), thermalization between the two phonon modes ($T_1 = T_2$) occurs (see Extended Data Fig. 1c, d). We also calculate the Casimir pressure caused by thermal fluctuations of the electromagnetic field (Extended Data Fig. 1e). In the distance range of our experiment ($d < 800$ nm), the thermal Casimir force (at 300 K) is less than 4% of the Casimir pressure driven by quantum fluctuations.

Device fabrication

The process of device fabrication is illustrated in Extended Data Fig. 5a–f. The process started with a silicon wafer (500 μm thick) coated with stoichiometric Si_3N_4 (100 nm thick) on both sides by low-pressure chemical vapour deposition. Photolithography was performed at the back side of the wafer and SF_6 plasma etching was used to remove the Si_3N_4 at the opening windows. The silicon wafer was then etched through in a KOH solution (25%, 80 $^\circ\text{C}$, 7 h), creating freestanding Si_3N_4 membranes at the front side of the wafer. For the right sample, the first photolithography and liftoff were performed to pattern contact electrodes (75 nm Au) on the surface of the membrane. The second photolithography and liftoff were performed to pattern spacers (150 nm Au). After that, the back side of the sample was evaporated with 75 nm Au. For the left sample, photolithography and SF_6 plasma etching were performed to define a square region ($500 \times 500 \mu\text{m}^2$) around the membrane, whose corners were designed to make contact with the spacers on the right sample. A short KOH (25%, 80 $^\circ\text{C}$, 15 min) etch was used to create a mesa structure with a depth of around 25 μm . After the KOH etch, the sample was evaporated with 75 nm Au on both sides.

After fabrication, the left sample was attached to a custom-made copper plate using conductive silver paint (Extended Data Fig. 5i) and the right sample was mounted on a printed circuit board (PCB) with the on-chip electrodes wire-bonded to the corresponding contact pads (Extended Data Fig. 5j).

Cleanliness of the membrane surfaces is crucial for parallel alignment of the two membranes at a short distance. We inspected samples under a confocal microscope, which can identify particles with sizes down to 100 nm. To maintain a high degree of cleanliness, we carried out the sample fabrication, wire-bonding and mounting of samples onto sample holders in a clean-room environment. Using atomic force microscopy, we measured the surface roughness of the membrane to be less than 1.5 nm.

Parallel alignment of the membrane samples

A diagram of the parallel alignment setup is shown in Extended Data Fig. 6a. The two samples were designed such that the corners of the mesa structure on the left sample align to the spacers on the right sample. The gold films coated on both sides of the membranes block the red detection laser ($\lambda = 633$ nm) while allowing dim transmission of blue illumination from a high-brightness LED ($\lambda = 460$ nm). A bias voltage, V_b , is applied between the two membrane surfaces. When the two samples are brought close and touch each other at the spacer regions, an electrical signal is picked up, pinpointing the corner that is touching. Simultaneous touching of the four corners indicates good alignment of parallelism. In the experiment, the distances at which the first corner touches and at which all four corners touch are within 80 nm.

Simultaneously, the parallelism between the membrane is monitored optically by imaging the brightness distribution of the interference pattern. Aligned membranes show uniform optical images when changing the distance (Extended Data Fig. 6b). By analysing the optical intensity at different locations of the membrane while changing the membrane separation, we obtain the relative tilting angles of the membranes with respect to the x and y axes as $\Delta\theta_x = 22 \pm 25 \mu\text{rad}$ and $\Delta\theta_y = 43 \pm 24 \mu\text{rad}$ (Extended Data Fig. 6d). For comparison, images of the misaligned membranes are shown in Extended Data Fig. 6c. In this case, $\Delta\theta_x = 228 \pm 33 \mu\text{rad}$ and $\Delta\theta_y = 179 \pm 39 \mu\text{rad}$ (Extended Data Fig. 6e).

Experimental setup and signal acquisition

The experimental setup is illustrated in Extended Data Fig. 7a. Two laser beams split from an intensity stabilized He/Ne laser (Thorlabs, HRS015B) are sent to the back sides of the membranes to interferometrically detect the thermal motion of the membranes. The distance between the mirror and the membrane is feedback-controlled by piezo-actuators, using the DC component of the reflected light (monitored by photodetectors) as a feedback signal. This is to achieve long-term stability of the detection sensitivity. The samples are mounted on a closed-loop thermoelectric heater/cooler, using a platinum resistance-temperature detector (RTD) as a temperature sensor connected to a temperature controller (Lakeshore 330). Feedback control of the bath temperatures allows stabilization and tuning of the mechanical frequencies. The right sample is mounted on a calibrated closed-loop piezo linear stage, which controls the relative distance between the membranes with a precision of around 5 nm.

The AC optical signals are detected by avalanched photodetectors and fed to lock-in amplifiers (Stanford Research, SR840). The lock-in reference frequency is set to be offset by 50 Hz from the mechanical resonance in order to prevent spiking of the reference signal. The time sequences of the two quadrature components are recorded by a data-acquisition system. Fast Fourier transform is then performed to obtain the thermal spectra of the mechanical modes. This method allows faster measurement of the thermal spectra compared with the frequency-sweeping benchtop spectrum analyser.

An optical image of the sample mount assembly and control stages is shown in Extended Data Fig. 7b. During the measurement, the whole assembly is kept in a vacuum chamber with pressure below 10^{-6} Torr. To avoid optical heating, laser powers are kept low at 8 μW before entering the chamber.

Electrostatic calibration of absolute distance between membranes

The relative distance between the membranes, d' , is controlled by a closed-loop piezo linear stage using a strain-gauge sensor. To calibrate the absolute distance between the membranes, $d = d' + d_0$, a widely used method in Casimir force measurements is to make use of the unique distance and voltage dependence of the electrostatic effect^{13,14,16}. In our experiment, the frequency splitting of the mechanical resonance due to electrostatic potentials follows $\Delta f_{\text{ele}} \propto K_E(d)[(V_b - V_0)^2 + V_{\text{rms}}^2]$, with electrostatic strength $K_E(d) \propto d^{-3}$. At each distance, the measured frequency splitting shows a parabolic dependence on the bias voltage (Fig. 4b). We fit the parabola curvatures (electrostatic strength $K_E(d)$) with a power law of $(d' + d_0)^{-3}$, using d_0 as the fitting parameter. Using this method, we determine the absolute distance between the membranes with a precision of around 5 nm (Extended Data Fig. 3a). From the fitting, we also obtain the surface potential, V_0 , at each distance (Extended Data Fig. 3b). The surface potential remains constant at different distances, agreeing with the theoretical prediction for a parallel-planes configuration³¹. (A distance dependence in V_0 is expected for the sphere-plane configuration in other Casimir force experiments.)

Thermal feedback control of mechanical frequencies

Without feedback control, the mechanical frequencies typically drift at a rate of around 0.3 Hz min^{-1} . During heat-transfer measurement, we apply feedback to control the bath temperatures and lock the mechanical resonance to a certain frequency. The frequency mismatch of the two modes can be maintained below 2 Hz, which is well below the linewidths of the two mechanical modes (4.6 Hz and 9.6 Hz) (Extended Data Fig. 8a). The time scale of the feedback loop is 16 s; the frequency is measured over 16 s and the heater/cooler power is adjusted at the end of this period.

Throughout the measurement, the sample bath temperatures are monitored by RTD sensors. Bath temperature fluctuations are unresolvable with the 0.01 K sensitivity of our temperature control system (Extended Data Fig. 8b). The bath temperature behaves similarly at all distances. This gives an upper bound for the temperature fluctuations of $\delta T_{\text{max}} = 0.01 \text{ K}$. To get a better estimation, we calculate δT on the basis of the fluctuations in the mechanical frequencies ($\delta f = 2 \text{ Hz}$; Extended Data Fig. 8a) and the measured frequency-temperature coefficient df/dT (2 kHz K^{-1}) of the membrane modes (Fig. 2d). The estimated fluctuation of the bath temperature δT is approximately 0.001 K.

Near-field thermal radiation effects

When the two membranes are brought close together, near-field thermal radiation could induce a deviation of the local temperatures on the sample surface from the temperature of the sample holder. This would lead to a difference between the actual and measured bath temperatures. At separations greater than 300 nm, the radiation heat transfer coefficient between gold surfaces has been measured³² to be less than $1.4 \text{ W m}^2 \text{ K}^{-1}$. Using the thermal conductivities of gold (150 W mK^{-1} at 75 nm)³⁷ and silicon nitride (10 W mK^{-1})³⁸ thin films, we calculate the local temperature deviation to be less than 0.02 K.

The local temperature change due to thermal radiation modifies the membrane stress and therefore induces a frequency downshift (or upshift) of mechanical mode 1 (or 2) when the membranes are close. This thermal radiation effect does not depend on the frequency matching of the two modes. To observe this effect, we first offset the frequencies of the two modes by 250 Hz by thermal tuning, and then fix the output of the heater and cooler (with feedback turned off). We note that the observed frequency shifts (Extended Data Fig. 4a) also include a contribution from the Casimir force (see Supplementary Information section 1). Such a shift is equal to half of the frequency splitting in the Casimir strong coupling regime (see Supplementary Information equation (S13)). We use the measured frequency splitting (Fig. 4c) to calculate the frequency shift caused by the Casimir force. For distances outside of the measured range, g_c is extrapolated using the power law $g_c \propto d^{-4.91}$. The corrected frequency shifts are less than 40 Hz when d is greater than 300 nm (Extended Data Fig. 4b). On the basis of the measured frequency-temperature coefficient of 2 kHz K^{-1} for the membrane modes (Fig. 2d), these frequency shifts correspond to

temperature changes of less than 0.02 K, which agrees with our calculation.

Stabilities of bias voltage and mechanical damping rates

Throughout the measurement, the bias voltage V_b is applied to compensate for the surface potential V_0 at each separation. V_b is sourced from a low-noise source meter (Keithley 2400) connected through an RC circuit, which serves as a potential divider and low-pass filter (see Supplementary Information section 3). We measured the noise spectral density of the source meter and estimated that its contribution to the noise of V_b reaches the thermal noise level at frequencies near the membrane resonance (182–194 kHz).

We characterize the mechanical damping rates of the two phonon modes ($\Delta\gamma_i/\bar{\gamma}_i$, where $i=1,2$) at different separations. The damping rates remain constant within the measurement error of $\pm 4\%$ in the whole distance range (Extended Data Fig. 9a). We also measure the damping rates when the bath temperatures are varied around the setpoints. With a temperature change of 0.3 K, the damping rates are constant within the measurement error of 4% (Extended Data Fig. 9b). Using the estimated bath temperature fluctuations ($\delta T = 0.001 \text{ K}$) obtained above, we estimate the temperature-induced fluctuations of mechanical damping to be less than 0.01%.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

36. Luo, Y., Zhao, R. K. & Pendry, J. B. van der Waals interactions at the nanoscale: the effects of nonlocality. *Proc. Natl Acad. Sci. USA* **111**, 18422–18427 (2014).
37. Langer, G., Hartmann, J. & Reichling, M. Thermal conductivity of thin metallic films measured by photothermal profile analysis. *Rev. Sci. Instrum.* **68**, 1510–1513 (1997).
38. Zhang, X. & Grigoropoulos, C. P. Thermal conductivity and diffusivity of free-standing silicon nitride thin films. *Rev. Sci. Instrum.* **66**, 1115–1120 (1995).

Acknowledgements The work was supported by the National Science Foundation (NSF) under grant number 1725335, the King Abdullah University of Science and Technology Office of Sponsored Research (OSR) (award numbers OSR-2016-CRG5-2950-03 and OSR-2016-CRG5-2996); and the Ernest S. Kuh Endowed Chair Professorship.

Author contributions R.Z., K.Y.F., H.-K.L. and X.Z. conceived the project. K.Y.F., H.-K.L. and R.Z. designed the experiment. K.Y.F. and H.-K.L. built the experimental setup, performed the measurement, and analysed the data. K.Y.F. and H.-K.L. fabricated the samples, with assistance from R.Z. and S.Y. R.Z. carried out numerical calculations of the Casimir force. K.Y.F., H.-K.L. and X.Z. wrote the manuscript with inputs from all authors. X.Z., Y.W. and S.Y. guided the research.

Competing interests The authors declare no competing interests.

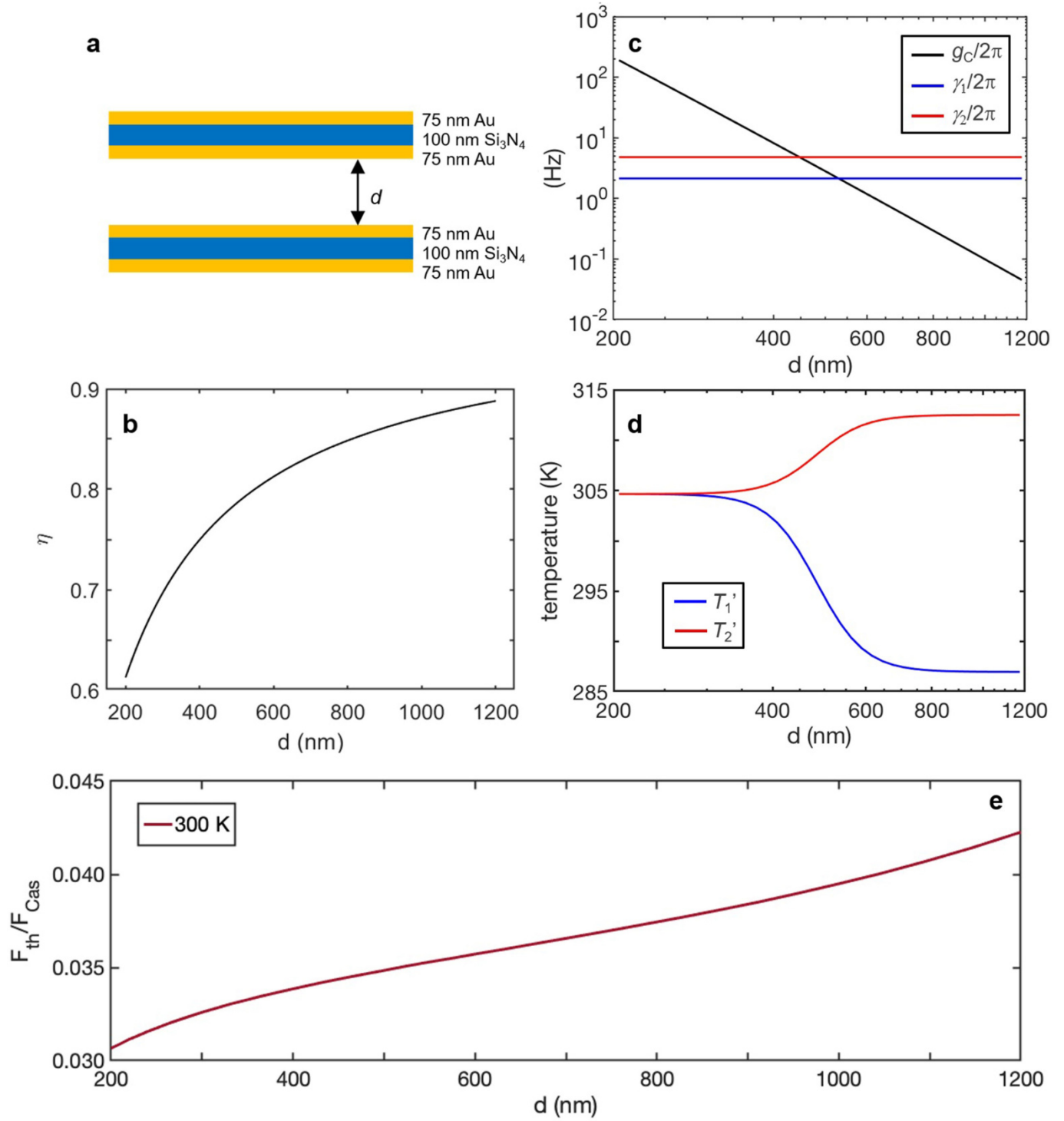
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1800-4>.

Correspondence and requests for materials should be addressed to X.Z.

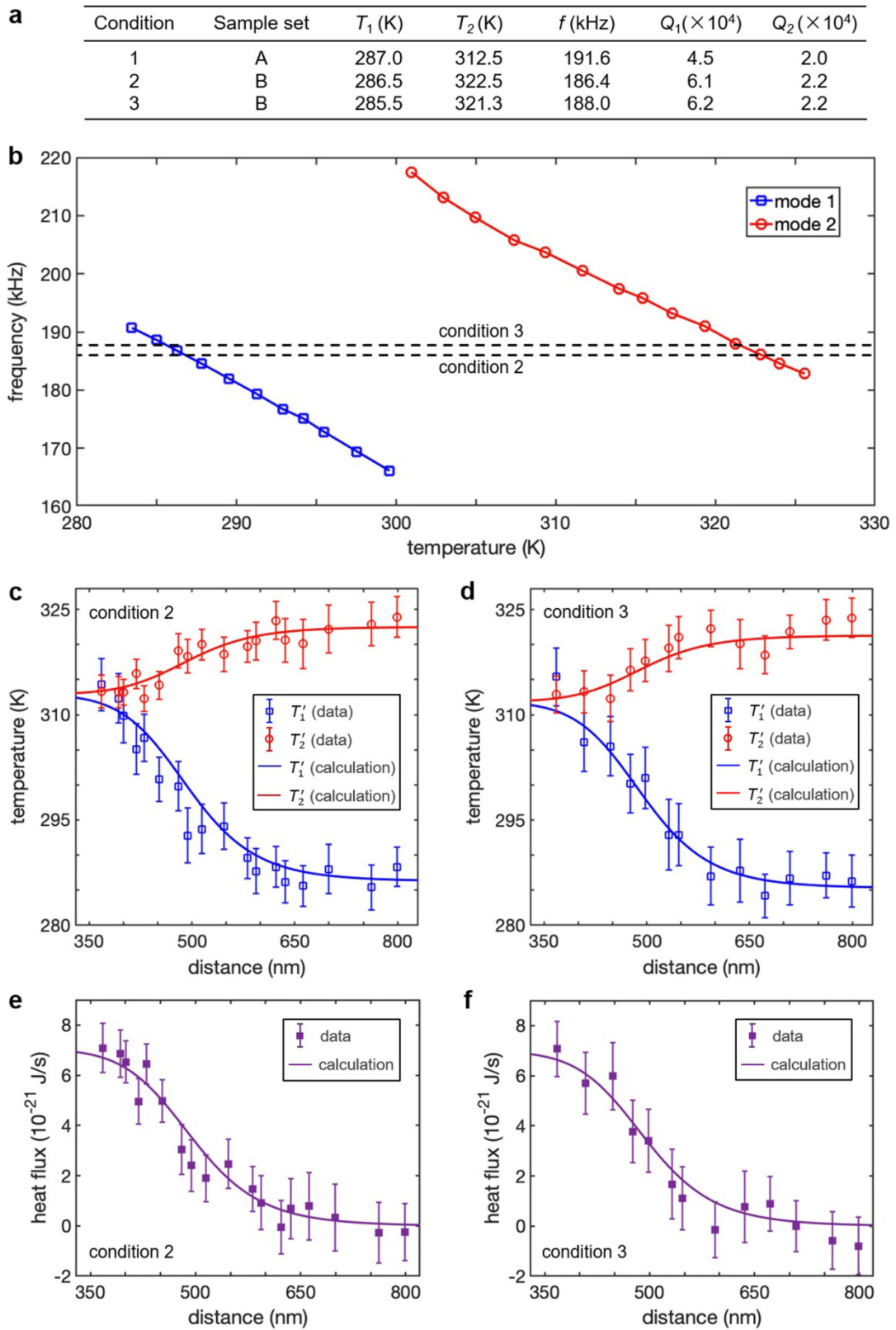
Peer review information Nature thanks Tal Carmon, Karthik Sasithirlu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



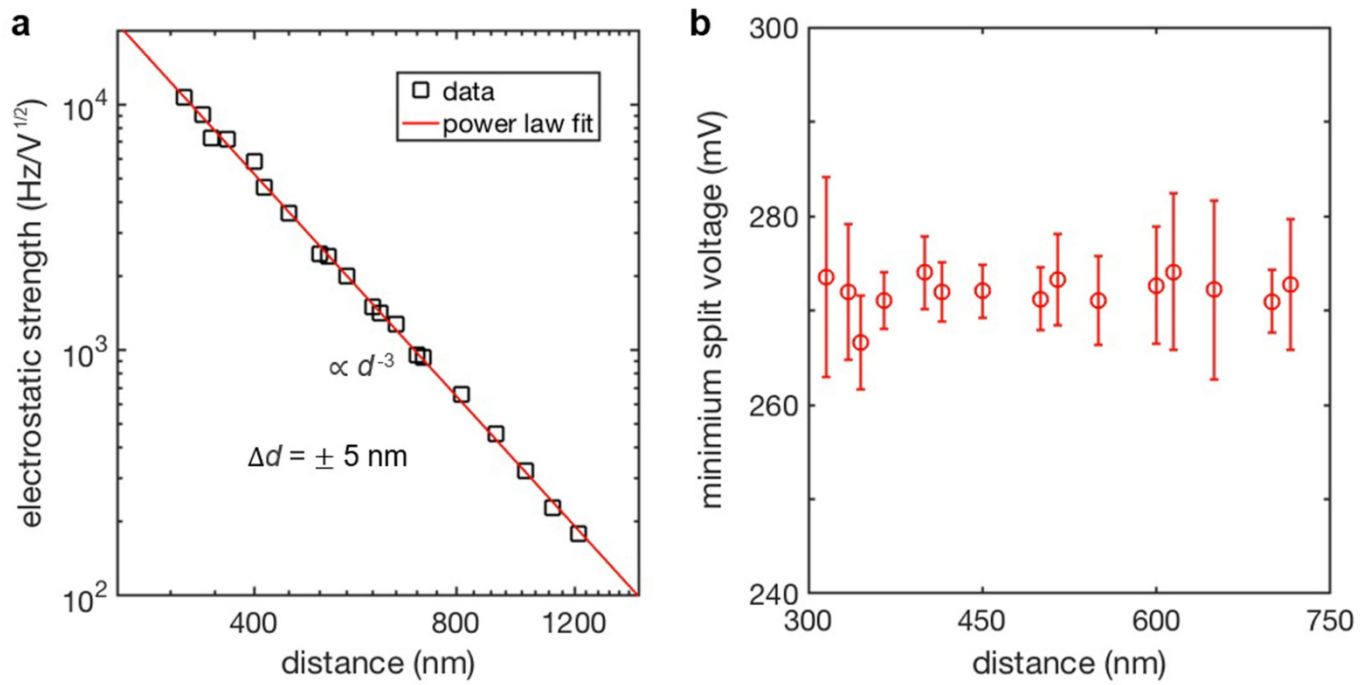
Extended Data Fig. 1 | Numerical calculations of the Casimir force and its heat transfer effect. **a**, Cross-section of the layered structure used in the experiment. **b**, Calculated correction factor, η , plotted against distance, d . **c**, Calculated coupling rate, g_C , plotted against d . **d**, Calculated mode

temperatures, T'_1 and T'_2 , plotted against d on the basis of experimental condition 1 in Extended Data Fig. 2a. **e**, Ratio between the Casimir pressures contributed from thermal fluctuations (F_{th}) and quantum vacuum fluctuations (F_{Cas}) plotted against d .

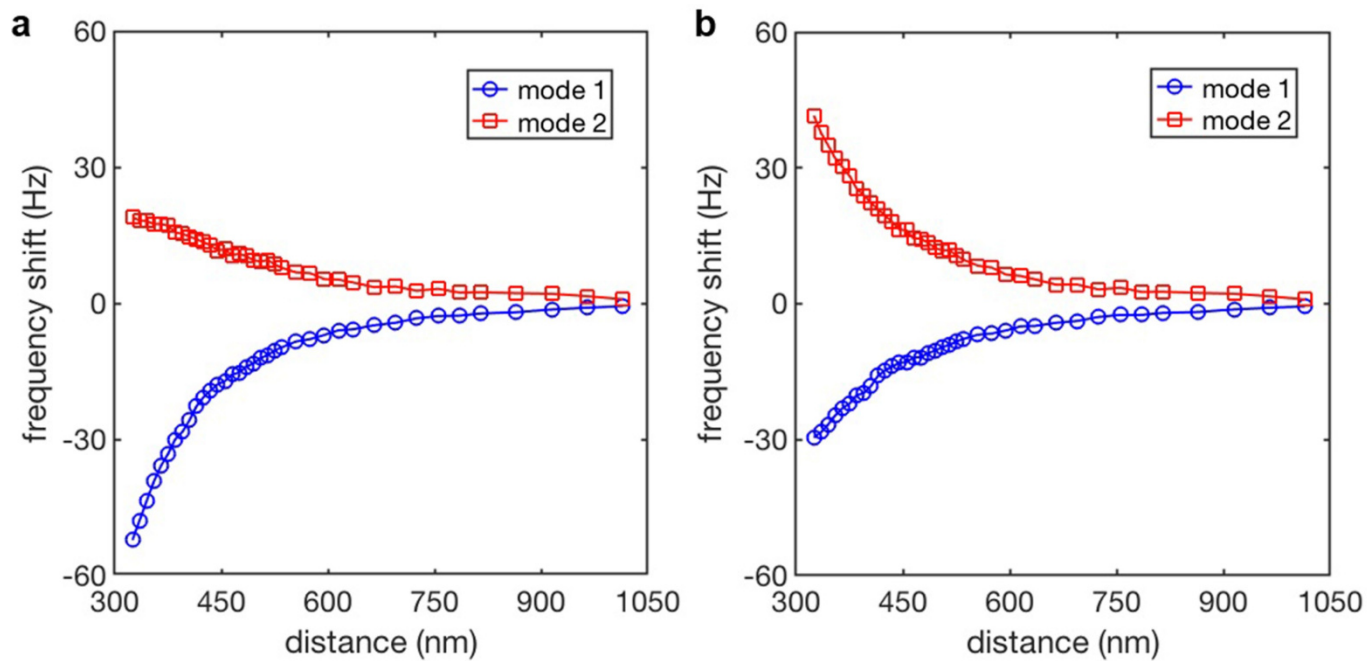


Extended Data Fig. 2 | Additional experimental results obtained from different samples and conditions. **a**, Summary of different experimental conditions used. Condition 1 corresponds to the experimental results presented in the main text. **b–f**, Measurement results obtained using conditions 2 and 3. In all cases, phonon mode splitting is examined and confirms that the Casimir force is dominant over the distance range concerned.

b, Resonance frequencies versus bath temperature for sample set B (conditions 2 and 3). **c, d**, Mode temperatures as functions of distances under different resonance-matching conditions. Error bars represent the standard error obtained from three hours of continuous measurement. **e, f**, Heat flux transferred across thermal baths as functions of distances. The error bars originate from error propagation in the calculation.

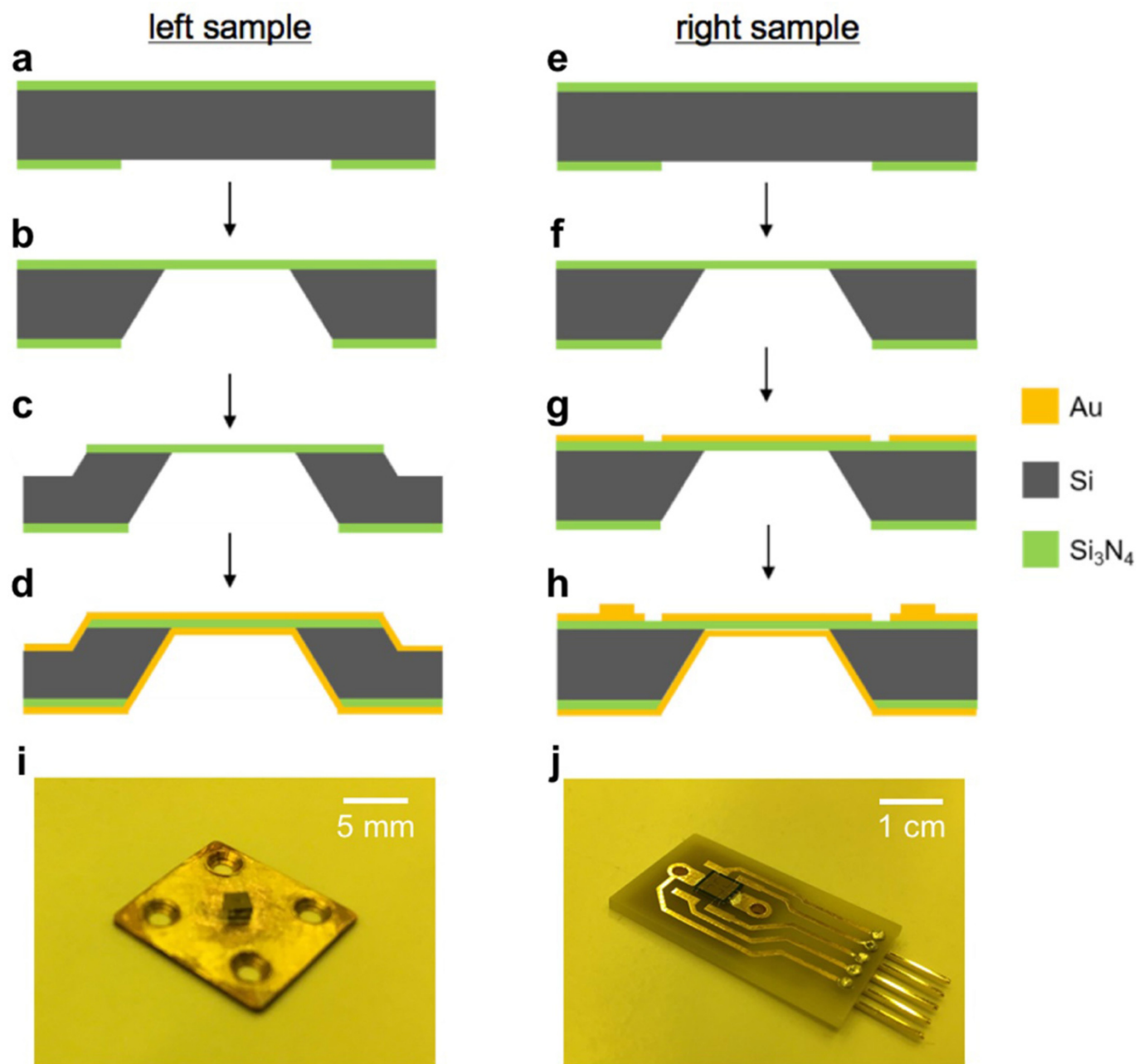


Extended Data Fig. 3 | Electrostatic calibration of the absolute distance between membranes. a, b, Dependence of electrostatic strength (a) and minimum splitting voltage V_0 (b) on the distance between membranes. In b, the error bars represent the error of the parabolic fit to the frequency splitting versus voltage.

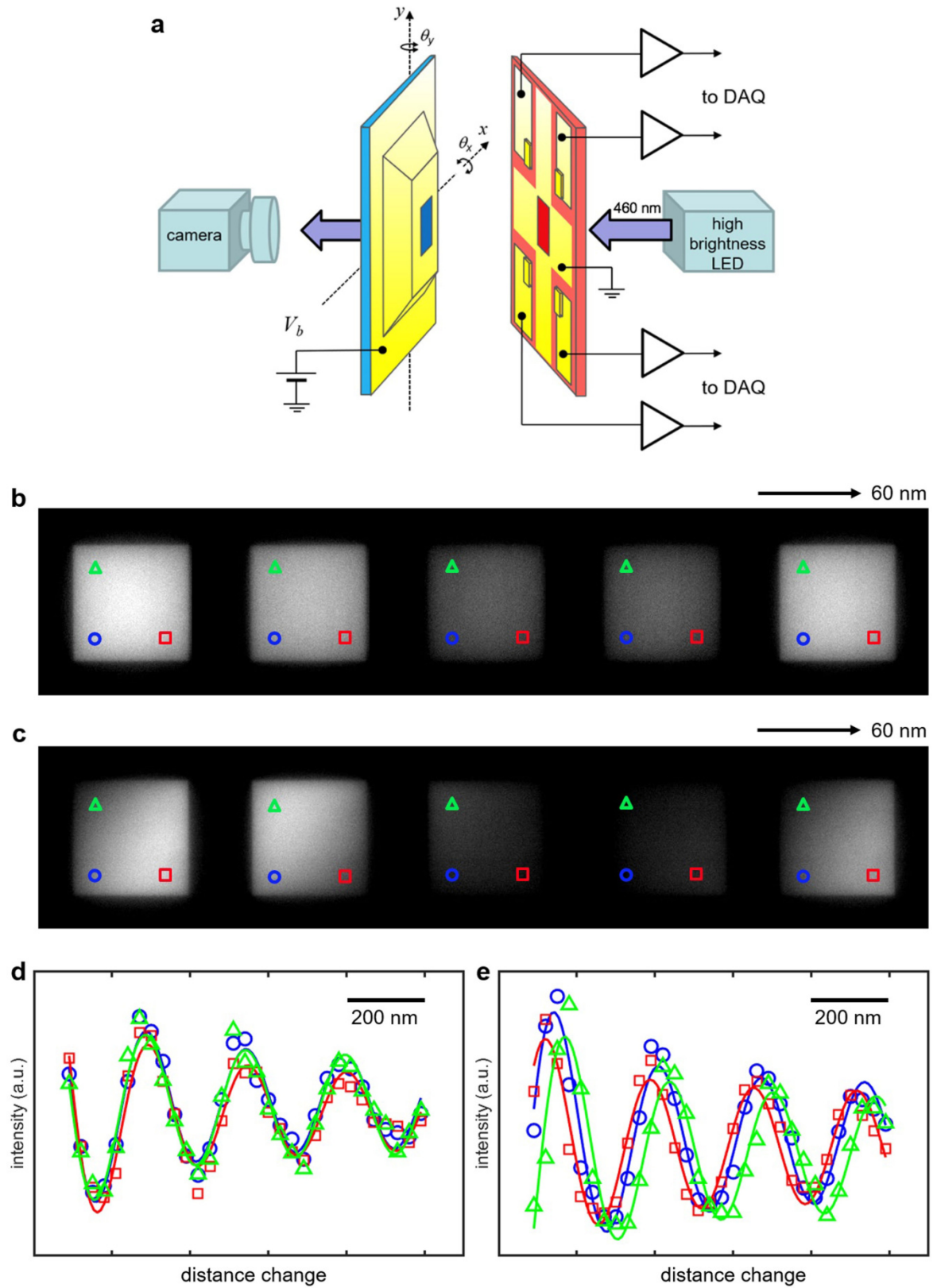


Extended Data Fig. 4 | Near-field thermal radiation effects. **a**, Frequency shifts of the two modes plotted against membrane distance. **b**, Frequency shifts of the two modes with the contribution from the Casimir force excluded.

Measurements were carried out at bath temperatures $T_1 = 287.0$ K and $T_2 = 312.5$ K. The frequencies of the modes are offset by 250 Hz.

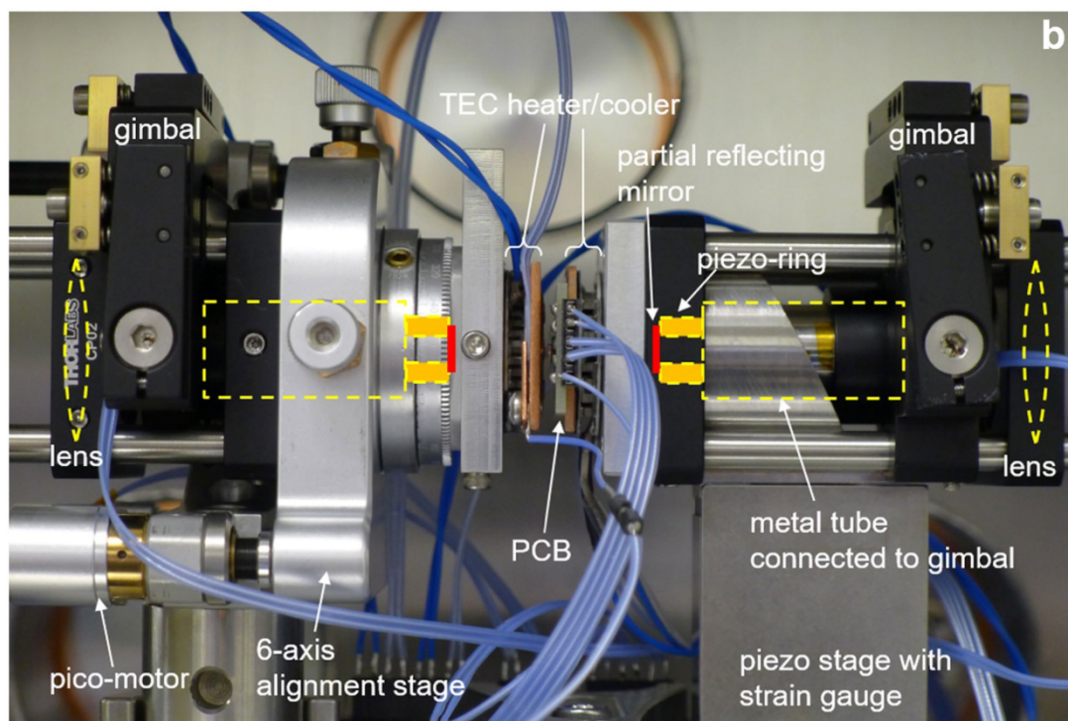
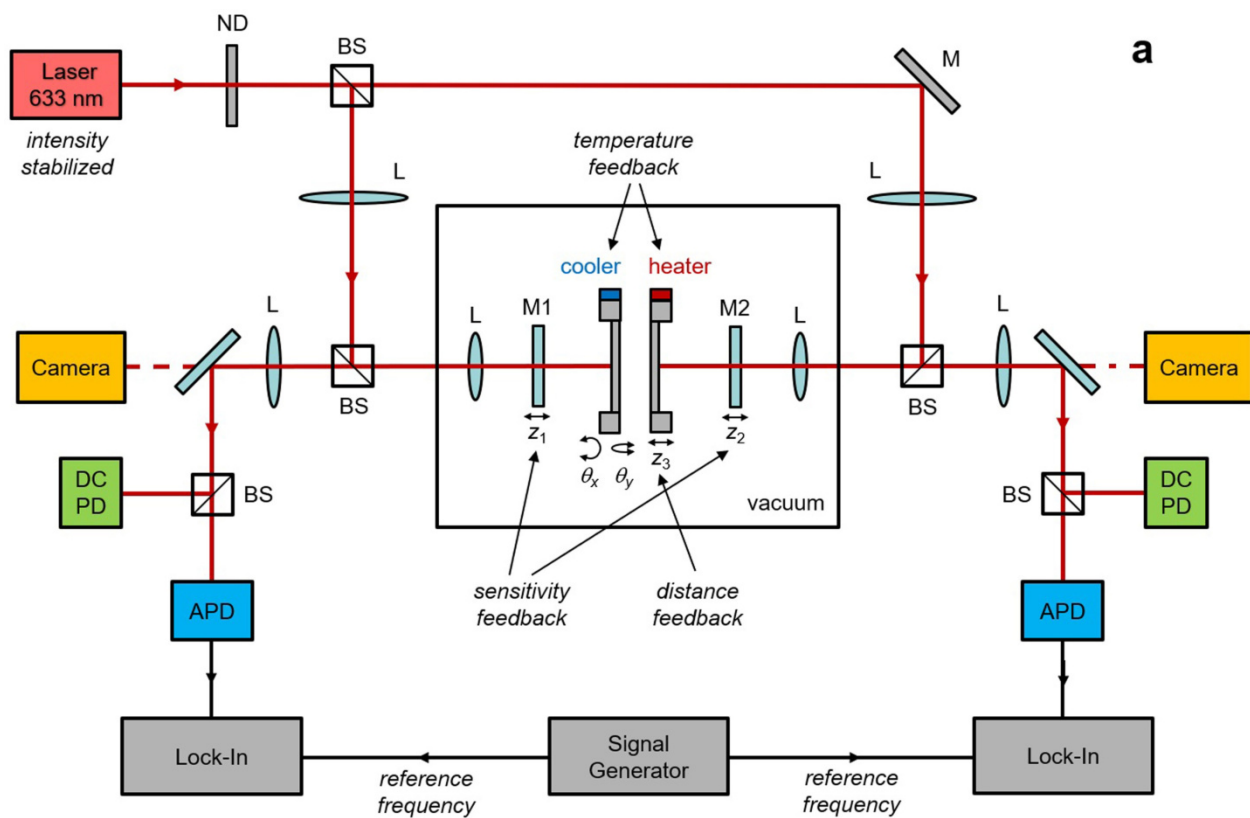


Extended Data Fig. 5 | Device fabrication. a–h, Fabrication process flow for the left (a–d) and right (e–h) samples. i, j, The left (i) and right (j) samples are attached to a custom-made copper plate and a printed circuit board, respectively.

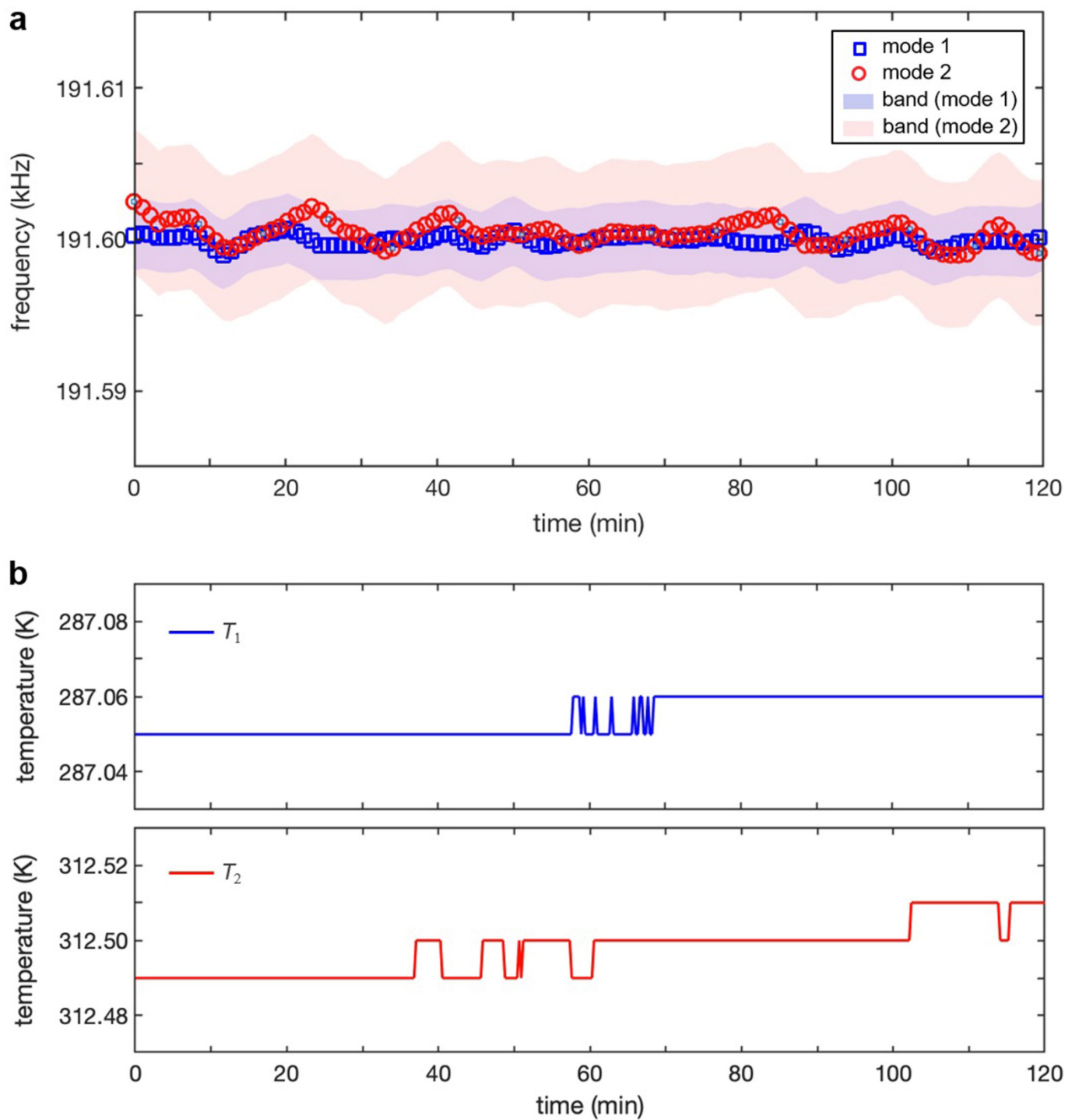


Extended Data Fig. 6 | Parallel alignment of the membranes. **a**, Schematic showing the parallel alignment setup. DAQ, data acquisition system. **b, c**, Transmission optical images of aligned (**b**) and misaligned (**c**) membranes. **d, e**, Optical intensity at different locations on the membranes (marked in **b, c**) as a function of the change in separation. Solid lines are sinusoidal fits with an

attenuation factor. The periodicity of around 230 nm matches well with the half-wavelength of the illumination light (460 nm). From the fitting, we find that the angle misalignments along the x and y directions are $\Delta\theta_x = 22 \pm 25 \mu\text{rad}$ and $\Delta\theta_y = 43 \pm 24 \mu\text{rad}$ for aligned membranes (**d**), and $\Delta\theta_x = 228 \pm 33 \mu\text{rad}$ and $\Delta\theta_y = 179 \pm 39 \mu\text{rad}$ for misaligned membranes (**e**).

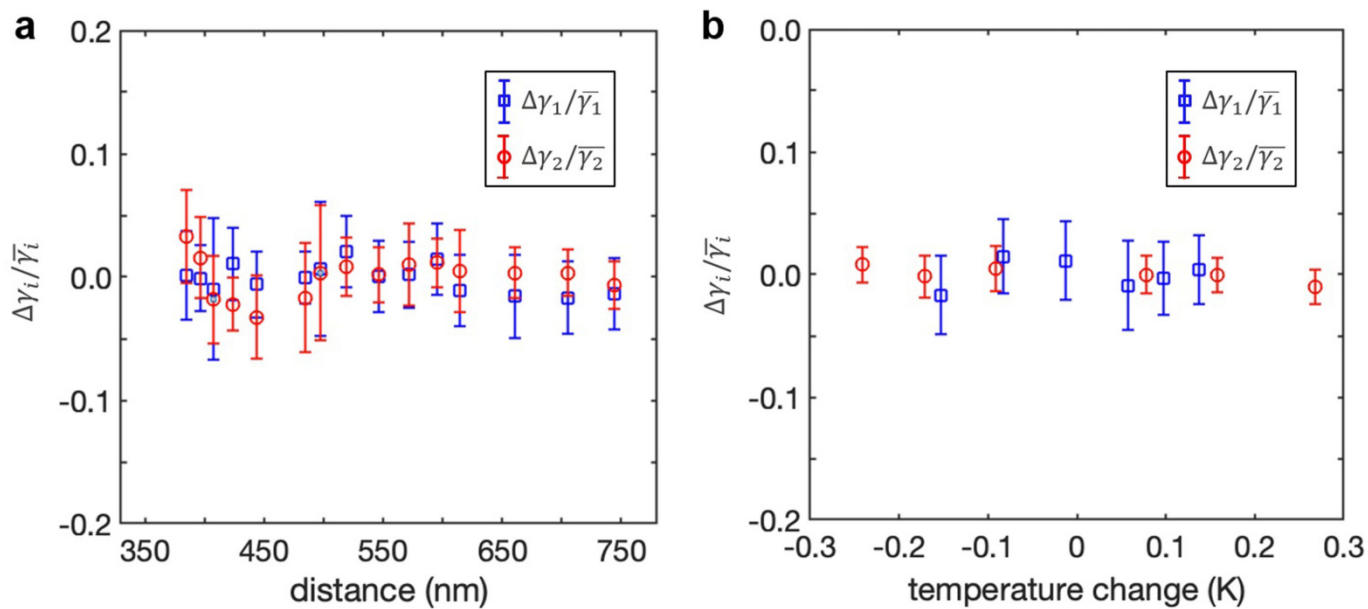


Extended Data Fig. 7 | Experimental setup. **a**, Schematic showing the experimental setup. APD, avalanche photodetector; BS, beam splitter; DC PD, DC photodetector; L, lens; M, mirror; ND, neutral density filter. **b**, Optical image of the sample mount assembly and control stages.



Extended Data Fig. 8 | Stability of mechanical frequency and temperature during thermal feedback. **a**, Frequency stability during thermal feedback control. The shaded areas represent the linewidths of the mechanical modes.

b, Bath temperatures read from the temperature controller during feedback control of the resonance frequencies.



Extended Data Fig. 9 | Characterization of mechanical damping rate. a, b, Relative change in damping rate ($\Delta\gamma_i/\bar{\gamma}_i$) plotted against distance (a) and temperature change (b). Error bars represent the standard deviation of 100 measurements.

A general theoretical and experimental framework for nanoscale electromagnetism

<https://doi.org/10.1038/s41586-019-1803-1>

Received: 4 February 2019

Accepted: 5 September 2019

Published online: 11 December 2019

Yi Yang^{1,6*}, Di Zhu^{1,6*}, Wei Yan^{2,3,4}, Akshay Agarwal¹, Mengjie Zheng^{1,5}, John D. Joannopoulos¹, Philippe Lalanne², Thomas Christensen^{1*}, Karl K. Berggren¹ & Marin Soljačić¹

The macroscopic electromagnetic boundary conditions, which have been established for over a century¹, are essential for the understanding of photonics at macroscopic length scales. Even state-of-the-art nanoplasmonic studies^{2–4}, exemplars of extremely interface-localized fields, rely on their validity. This classical description, however, neglects the intrinsic electronic length scales (of the order of ångström) associated with interfaces, leading to considerable discrepancies between classical predictions and experimental observations in systems with deeply nanoscale feature sizes, which are typically evident below about 10 to 20 nanometres^{5–10}. The onset of these discrepancies has a mesoscopic character: it lies between the granular microscopic (electronic-scale) and continuous macroscopic (wavelength-scale) domains. Existing top-down phenomenological approaches deal only with individual aspects of these omissions, such as nonlocality^{11–13} and local-response spill-out^{14,15}. Alternatively, bottom-up first-principles approaches—for example, time-dependent density functional theory^{16,17}—are severely constrained by computational demands and thus become impractical for multiscale problems. Consequently, a general and unified framework for nanoscale electromagnetism remains absent. Here we introduce and experimentally demonstrate such a framework—amenable to both analytics and numerics, and applicable to multiscale problems—that reintroduces the electronic length scale via surface-response functions known as Feibelman d parameters^{18,19}. We establish an experimental procedure to measure these complex dispersive surface-response functions, using quasi-normal-mode perturbation theory and observations of pronounced nonclassical effects. We observe nonclassical spectral shifts in excess of 30 per cent and the breakdown of Kreibig-like broadening in a quintessential multiscale architecture: film-coupled nanoresonators, with feature sizes comparable to both the wavelength and the electronic length scale. Our results provide a general framework for modelling and understanding nanoscale (that is, all relevant length scales above about 1 nanometre) electromagnetic phenomena.

We reintroduce the electronic length scale by amending the classical boundary conditions (Fig. 1a) with a set of mesoscopic complex surface-response functions, known as the Feibelman d_{\perp} and d_{\parallel} parameters (Fig. 1b)^{18,19}, which play a role analogous to the local bulk permittivity, but for interfaces between two materials. d_{\perp} and d_{\parallel} are equal to the frequency-dependent centroids of the induced charge and the normal derivative of the tangential current, respectively, at an equivalent planar interface (Fig. 1c and Supplementary Information section S1). They enable a leading-order-accurate incorporation of nonlocality, spill-out and surface-enabled Landau damping (tunnelling¹⁴ and size quantization²⁰, which are not incorporated in the d parameters, are non-negligible at feature sizes below about 1 nm; see Supplementary Information section S2.C).

We start by summarizing the key elements of our framework. The d parameters drive an effective nonclassical surface polarization $\mathbf{P}_s \equiv \boldsymbol{\pi} + i\omega^{-1}\mathbf{K}$ (Fig. 1d, Supplementary Information section S2.A)^{21,22}, with d_{\perp} contributing an out-of-plane surface dipole density of $\boldsymbol{\pi} \equiv d_{\perp}\epsilon_0[\mathbf{E}_{\perp}]\hat{\mathbf{n}}$ (ϵ_0 is the vacuum permittivity) and d_{\parallel} contributing an in-plane surface current density of $\mathbf{K} \equiv i\omega d_{\parallel}[\mathbf{D}_{\parallel}]$. Here, $[\mathbf{f}] \equiv \mathbf{f}^+ - \mathbf{f}^-$ denotes the discontinuity of a field \mathbf{f} (an electric \mathbf{E} , magnetic \mathbf{B} , displacement \mathbf{D} or magnetizing \mathbf{H} field oscillating at frequency ω) across an interface $\partial\Omega$ with outward normal $\hat{\mathbf{n}}$ (the superscripts ‘+’ and ‘−’ denote points immediately outside and inside the interface, that is, $\mathbf{f}^{\pm} \equiv \mathbf{f}(\mathbf{r} \pm 0^{\pm}\hat{\mathbf{n}})$, where $\mathbf{r} \in \partial\Omega$). Similarly, $f_{\perp} \equiv \hat{\mathbf{n}} \cdot \mathbf{f}$ and $\mathbf{f}_{\parallel} \equiv (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^T)\mathbf{f}$ denote the (scalar) perpendicular and (vectorial) parallel components of \mathbf{f} relative to $\partial\Omega$. These surface terms can be equivalently incorporated

¹Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Laboratoire Photonique Numérique et Nanosciences, Institut d’Optique d’Aquitaine, Université Bordeaux, CNRS, Talence, France. ³School of Engineering, Westlake University, Hangzhou, China. ⁴Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, China. ⁵College of Mechanical and Vehicle Engineering, State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha, China. ⁶These authors contributed equally: Yi Yang, Di Zhu. *e-mail: yiy@mit.edu; dizhu@mit.edu; tchr@mit.edu

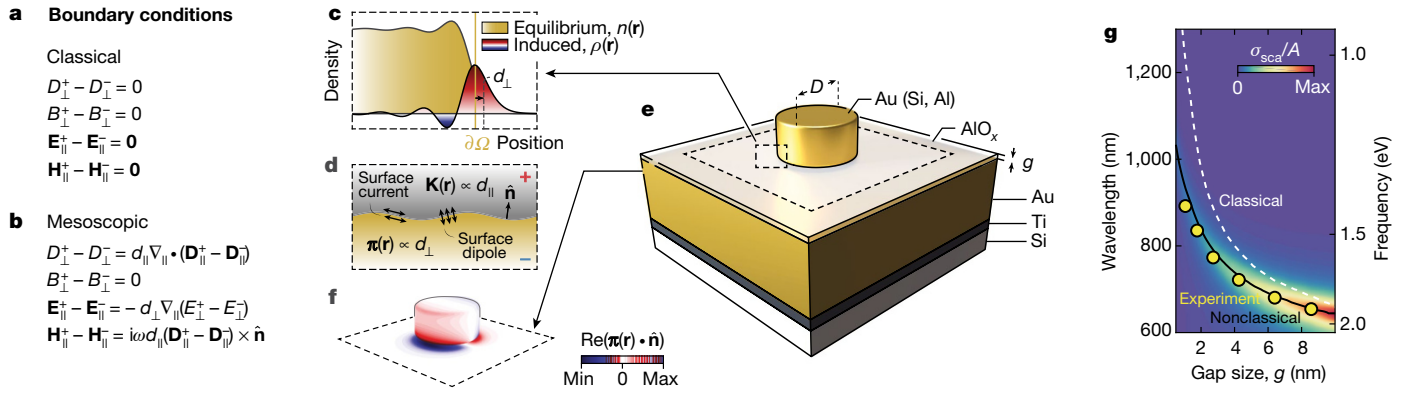


Fig. 1 | Framework, experimental structure and measured nonclassical shifts. **a**, **b**, Classical (**a**) and mesoscopic (**b**) electromagnetic boundary conditions. **c**, Equilibrium and induced densities, $n(\mathbf{r})$ and $\rho(\mathbf{r}, \omega)$ (not to scale) at a jellium–vacuum interface (Wigner–Seitz radius, $r_s = 3.93$ atomic units; $\hbar\omega = 1 \text{ eV}$; \hbar , reduced Planck constant) computed from (time-dependent) density functional theory. d_{\perp} is the centroid of the induced charge. **d**, Nonclassical corrections can be formulated as self-consistent surface polarizations, representing the effective surface dipole density $\boldsymbol{\pi}(\mathbf{r})$ and the current density $\mathbf{K}(\mathbf{r})$. **e**, Experimental structure: film-coupled Au nanodisks on a

as a set of mesoscopic boundary conditions (here without external interface currents or charges) for the conventional macroscopic Maxwell equations (also shown in Fig. 1b; see Supplementary Information section S2.B)

$$[D_{\perp}] = -i\omega^{-1} \nabla_{\parallel} \cdot \mathbf{K} = d_{\perp} \nabla_{\parallel} \cdot [\mathbf{D}_{\parallel}] \quad (1a)$$

$$[B_{\perp}] = 0 \quad (1b)$$

$$[\mathbf{E}_{\parallel}] = -\varepsilon_0^{-1} \nabla_{\parallel} \pi = -d_{\perp} \nabla_{\parallel} [E_{\perp}] \quad (1c)$$

$$[\mathbf{H}_{\parallel}] = \mathbf{K} \times \hat{\mathbf{n}} = i\omega d_{\parallel} [\mathbf{D}_{\parallel}] \times \hat{\mathbf{n}} \quad (1d)$$

These mesoscopic boundary conditions are a twofold generalization from opposite directions. First, they generalize the usual macroscopic electromagnetic boundary conditions ($[D_{\perp}] = [B_{\perp}] = 0$ and $[\mathbf{E}_{\parallel}] = [\mathbf{H}_{\parallel}] = \mathbf{0}$) to which they reduce in the limit $d_{\perp} = d_{\parallel} = 0$. Secondly, they represent a conceptual and practical generalization of the applicability of the Feibelman d parameters, elevated from their original purview of planar¹⁸ and spherical²³ interfaces, and beyond recent quasistatic considerations²², to a fully general electrodynamic framework.

Experimentally, we establish a systematic approach to measure the d -parameter dispersion of a general two-material interface and illustrate it using Au– AlO_x interfaces. Whereas the d parameters of simple metals can be accurately computed within jellium time-dependent density functional theory^{18,24}, those of noble metals, such as Au, require time-dependent density functional theory beyond the jellium approximation owing to non-negligible screening from lower-lying orbitals^{22,25,26}. We show that the d parameters can instead be measured experimentally. By developing and exploiting a quasi-normal-mode (QNM)-based^{27,28} perturbation expression, we translate these mesoscopic quantities directly into observables—spectral shifting and broadening—and measure them in designed plasmonic systems that exhibit pronounced nonclassical corrections. Our experimental testbed enables a direct procedure to extract the d parameters from standard dark-field spectroscopic measurements, in a manner analogous to ellipsometric measurements of the local bulk permittivity. Moreover, by investigating a complementary hybrid plasmonic setup, we discover and experimentally demonstrate design principles for structures that are classically robust—that is, they exhibit minimal nonclassical corrections—even under nanoscopic conditions.

The extensive interest in film-coupled nanoresonators^{3,4}, which combine wavelength-scale resonators with nanometric gaps, is a particularly pertinent example that underscores the need for

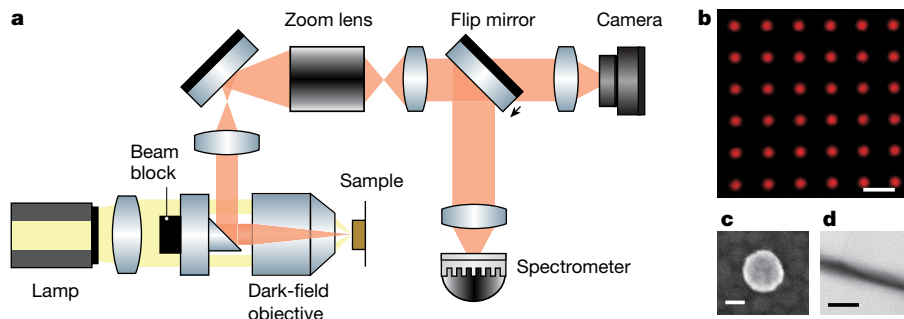


Fig. 2 | Measurement setup and sample micrographs. **a**, Tabletop dark-field scattering setup, where dark-field scattering from nanodisks illuminated by a broadband lamp is collected by a camera and a spectrometer (for details see Methods and Supplementary Information section S10). **b**, Dark-field optical

micrograph of a Au nanodisk array (scale bar, 2 μm). **c**, Scanning electron microscopy image of a single Au nanodisk (scale bar, 40 nm). **d**, Cross-sectional transmission electron microscopy image of an AlO_x gap (scale bar, 10 nm).

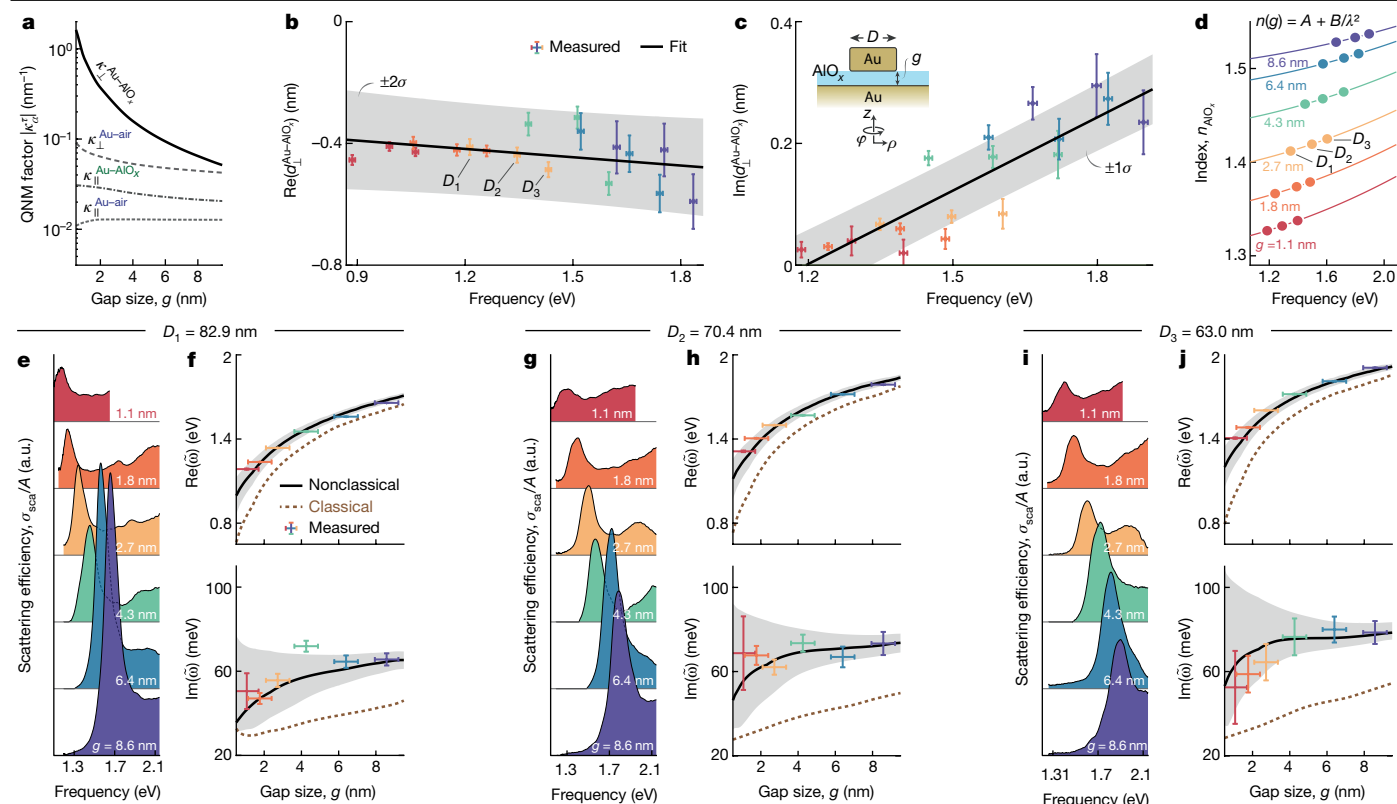


Fig. 3 | Systematic measurement of the complex surface-response function $d_{\perp}(\omega)$ of the Au-AIO_x interface. **a**, Nonclassical perturbation strengths, calculated from QNM-based perturbation theory (equation (6)) in a film-coupled Au nanodisk (see inset of schematic in **c**, $D = 70.4$ nm); $\kappa_{\perp}^{\text{Au-AIO}_x}$ is dominant. **b**, **c**, Measured (markers) dispersion of $\text{Re}(d_{\perp}^{\text{Au-AIO}_x})$ (**b**) and $\text{Im}(d_{\perp}^{\text{Au-AIO}_x})$ (**c**) and their linear fits (lines). Gap sizes g are distinguished by colour, and diameters ($D_1 \approx 82.9$ nm, $D_2 \approx 70.4$ nm and $D_3 \approx 63.0$ nm) decrease to the right. Inset to **c**, sample structure schematic. **d**, Measured thickness-dependent refractive indices of bare AIO_x films grown on Au for diameters D_1 ,

D_2 and D_3 . **e–j**, Scattering efficiency (**e**, **g**, **i**) across distinct diameters and gap sizes and extracted complex (1, 1) resonance eigenfrequencies (markers; **f**, **h**, **j**). Whereas classical predictions (brown dashed lines) deviate considerably from the observations, our nonclassical calculations (black solid lines), employing the aforementioned linear $d_{\perp}^{\text{Au-AIO}_x}(\omega)$ fit, are in quantitative agreement across all diameters. Shadings indicate fit-derived confidence intervals for our calculations: $2\sigma \approx 95\%$ for $\text{Re}(d_{\perp}^{\text{Au-AIO}_x})$ and $\text{Re}(\tilde{\omega})$ (**b**, **f**, **h**, **j**) and $1\sigma \approx 68\%$ for $\text{Im}(d_{\perp}^{\text{Au-AIO}_x})$ and $\text{Im}(\tilde{\omega})$ (**c**, **f**, **h**, **j**). a.u., arbitrary units. In **e–j**, the diameter D of the Au sample used is shown above the panels.

multiscale electrodynamic tools that incorporate nonclassical effects. We designed and fabricated film-coupled nanodisks (Figs. 1e, 2b–d) of various materials to verify our framework and directly measure the d parameters. Specifically, an optically thick Au film (atop a Si substrate) was separated from lithographically defined Au, Si or Al nanodisks (of diameter D) by a nanoscale AIO_x spacer, deposited by atomic layer deposition (ALD; see Supplementary Information section S5), demarcating a film–nanodisk gap of thickness g . The nanodisks support localized gap–plasmon resonances, which are (p, q) integer-indexable according to their field variations in the azimuthal and radial directions, respectively²⁹. The fundamental mode (1, 1) is optically accessible in the far field and exhibits highly confined electromagnetic fields within the gap, suggesting potentially large nonclassical corrections.

We implemented the mesoscopic boundary conditions (Fig. 1b) in a standard full-wave numerical solver (COMSOL Multiphysics, implementation available at <https://github.com/yiy-mit/nanoEM>; see Supplementary Information section S3). With specified d parameters, this permits self-consistent calculations of, for example, the nonclassical surface dipole density $\mathbf{\pi}(\mathbf{r})$, as shown in Fig. 1f for the (1, 1) mode. Similarly, conventional electromagnetic quantities such as the scattering efficiency σ_{scat}/A ($A = \pi D^2/4$) can be computed, enabling comparison with experimental results (Fig. 1g). For Au disks, the (1, 1) resonance is consistently blueshifted relative to the classical prediction, with shifts exceeding 30% for the smallest considered gaps.

To extract the surface-response functions from observables, we develop a perturbative description³⁰ of the nonclassical spectral shift using the solver QNMEig²⁸ of the QNM freeware MAN (Modal Analysis of Nanoresonators), with retardation incorporated explicitly: the true eigenfrequency $\tilde{\omega} = \tilde{\omega}^{(0)} + \tilde{\omega}^{(1)} + \dots$ (eigenindex implicit) has a first-order nonclassical correction $\tilde{\omega}^{(1)}$ to its classical value $\tilde{\omega}^{(0)}$ (Supplementary Information section S4)

$$\tilde{\omega}^{(1)} = \tilde{\omega}^{(0)} \sum_{\tau} \kappa_{\perp}^{\tau} d_{\perp}^{\tau} + \kappa_{\parallel}^{\tau} d_{\parallel}^{\tau} \quad (2)$$

with mode-, shape- and scale-dependent nonclassical perturbation strengths (units of inverse length)

$$\kappa_{\perp}^{\tau} \equiv - \int_{\partial\Omega^{\tau}} \tilde{D}_{\perp}^{(0)} [\tilde{E}_{\perp}^{(0)}] d^2\mathbf{r} \quad \text{and} \quad \kappa_{\parallel}^{\tau} \equiv \int_{\partial\Omega^{\tau}} \tilde{E}_{\parallel}^{(0)} \cdot [\tilde{D}_{\parallel}^{(0)}] d^2\mathbf{r} \quad (3)$$

Here, τ runs over all material interfaces such that $\bigcup_{\tau} \partial\Omega^{\tau} = \partial\Omega$, that is, $\tau \in \{\text{Au-AIO}_x, \text{Au-air}\}$ for our setup, and $\tilde{\mathbf{D}}^{(0)}$ and $\tilde{\mathbf{E}}^{(0)}$ denote the \mathbf{D} and \mathbf{E} fields of the (suitably normalized; see Supplementary Information section S4.A) classical QNM under consideration. Conceptually, equation (3) states that the nonclassical perturbation strength is proportional to the difference between the classical surface energy densities evaluated on either side of the interface.

Figure 3a shows the magnitudes of the nonclassical perturbation strengths in a film-coupled Au nanodisk: $\kappa_{\perp}^{\text{Au-AIO}_x}$, $\kappa_{\perp}^{\text{Au-air}}$, $\kappa_{\parallel}^{\text{Au-AIO}_x}$ and $\kappa_{\parallel}^{\text{Au-air}}$. Evidently, $\kappa_{\perp}^{\text{Au-AIO}_x}$ far exceeds $\kappa_{\parallel}^{\text{Au-AIO}_x}$ for all gap sizes of interest, rendering

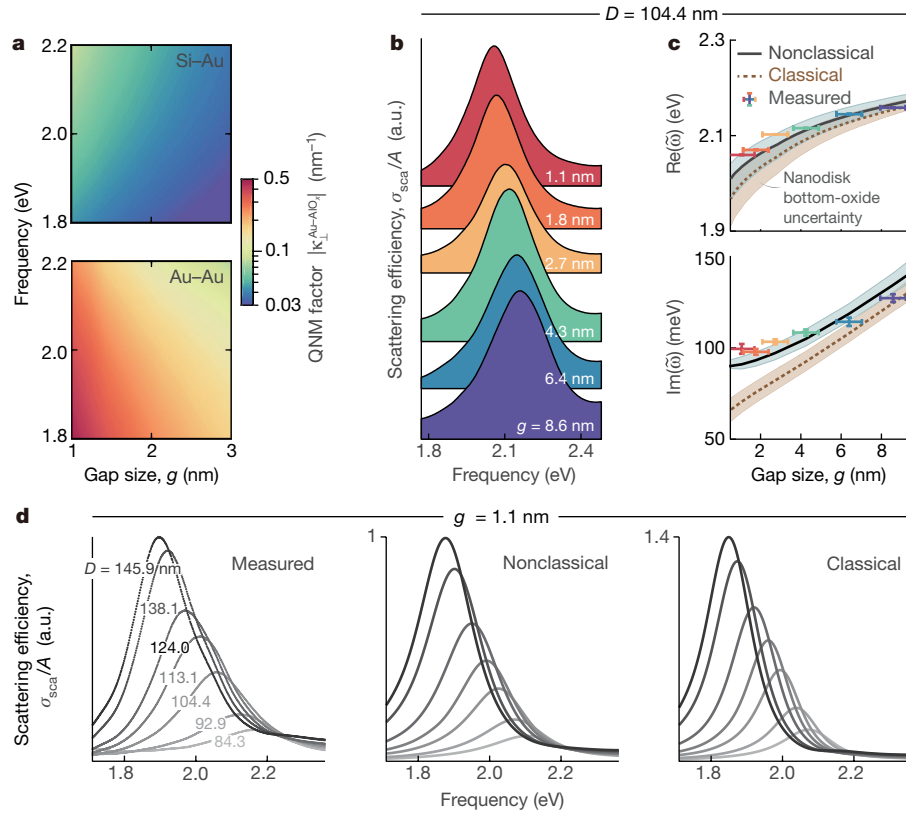


Fig. 4 | Robustness to nonclassical corrections. **a**, The nonclassical perturbation strength is one order of magnitude smaller in the hybrid Si-Au system than in its Au-Au counterpart. The Si and Au nanodisk diameters are chosen to ensure spectral alignment of the (1, 1) resonance at every gap size (spanning $D \in [80 \text{ nm}, 160 \text{ nm}]$ and $D \in [15 \text{ nm}, 40 \text{ nm}]$, respectively). **b–d**, Observation of robust optical response in a Si-Au setup with the

detrimental quantum corrections mitigated. The nonclassical calculation for the Si-Au setup assumes $d_{\perp}^{\text{Au-AuO}_x} = -0.5 + 0.3i$ nm, a constant extrapolation to higher frequencies from Fig. 3b, c. **In d**, the measured and calculated spectra are normalized separately. The calculated spectra incorporate inhomogeneous broadening (about 6%) due to disk-size inhomogeneity (Supplementary Information sections S7, S12.B).

the impact of $d_{\perp}^{\text{Au-AuO}_x}$ negligible. Similarly, the impact of $d_{\perp}^{\text{Au-air}}$ is negligible relative to $d_{\perp}^{\text{Au-AuO}_x}$ because $\kappa_{\perp}^{\text{Au-AuO}_x} \gg \kappa_{\perp}^{\text{Au-air}}$ for small g , and more generally because $d_{\perp}^{\text{Au-AuO}_x} > d_{\perp}^{\text{Au-air}}$ owing to screening from AlO_x (Supplementary Information section S16)³¹. Jointly, this justifies the approximation

$$\tilde{\omega}^{(1)} \approx \tilde{\omega}^{(0)} \kappa_{\perp}^{\text{Au-AuO}_x} d_{\perp}^{\text{Au-AuO}_x} \quad (4)$$

Inversion of equation (4) enables the direct experimental inference of $d_{\perp}^{\text{Au-AuO}_x}$, given the measured $\tilde{\omega}$ and calculated $\tilde{\omega}^{(0)}$ (because, to first order, $\tilde{\omega}^{(1)} \approx \tilde{\omega} - \tilde{\omega}^{(0)}$). We note that $|\text{Re}(\kappa_{\perp}^{\text{Au-AuO}_x})| \gg |\text{Im}(\kappa_{\perp}^{\text{Au-AuO}_x})|$ (by 1–2 orders of magnitude) for the considered gap sizes; consequently, $\text{Re}(d_{\perp}^{\text{Au-AuO}_x})$ contributes to spectral shifting and $\text{Im}(d_{\perp}^{\text{Au-AuO}_x})$ to broadening.

Ensemble scattering spectra of 18 arrays of Au nanodisks (height, 31 nm) were measured by dark-field scattering microscopy (Fig. 2, Methods), spanning three diameters and six gap sizes (Fig. 3e–j). The associated complex eigenfrequencies $\tilde{\omega}$ were subsequently extracted by Voigt profile deconvolution (Supplementary Information section S12), using the measured particle size distribution (Supplementary Information section S7). For the AlO_x spacer, we observed ellipsometrically (and include in our calculations) a thickness-dependent refractive index n_{AlO_x} (Fig. 3d, Supplementary Information section S9), an effect commonly observed in ultrathin ALD-grown AlO_x layers³² and other ALD-grown materials³³.

Figure 3b, c shows the complex surface-response function $d_{\perp}^{\text{Au-AuO}_x}(\omega)$, extracted using equation (4). Within the considered spectral range, $\text{Re}(d_{\perp}^{\text{Au-AuO}_x})$ (Fig. 3b) reveals a nearly dispersionless surface response of comparatively large magnitude, from -0.5 nm to -0.4 nm. By contrast, $\text{Im}(d_{\perp}^{\text{Au-AuO}_x})$ (Fig. 3c) is strongly dispersive, increasing from $\lesssim 0.1$ nm in

the near-infrared to about 0.3 nm in the visible. The thickness dependence of n_{AlO_x} imparts an attendant implicit dependence on the inferred $d_{\perp}^{\text{Au-AuO}_x}(\omega)$ (Supplementary Information section S13). As a result, the frequency fits in Fig. 3b, c convey a composite dependence along the $(\omega, n_{\text{AlO}_x})$ space (Fig. 3d, circles) sampled by our data.

Whereas the negative sign of $\text{Re}(d_{\perp}^{\text{Au-AuO}_x})$, and the associated blueshift of $\text{Re}(\tilde{\omega} - \tilde{\omega}^{(0)})$ (Fig. 3f, h, j, top panels), agrees with earlier observations in Au (refs. 7,8) and Ag (refs. 6,9,10) nanoparticles, the spectral shift is substantially larger. There are two reasons for this: first, the nonclassical perturbation strength $\kappa_{\perp}^{\text{Au-AuO}_x}$ is much larger than, for example, in standalone nanospheres or film-coupled nanospheres, owing to the strong field confinement beneath the entire nanodisk footprint (Supplementary Information section S4); second, screening from the AlO_x cladding expels induced charge into Au, thereby enhancing $d_{\perp}^{\text{Au-AuO}_x}$ relative to the unscreened interface, that is, relative to $d_{\perp}^{\text{Au-air}}$ (Supplementary Information section S16)³¹.

Nonclassical broadening due to surface-enhanced Landau damping, that is, $\text{Im}(\tilde{\omega} - \tilde{\omega}^{(0)})$, is similarly enhanced for the same reasons (Fig. 3f, h, j, lower panels). Classically, the linewidth reduces nearly monotonically with gap size, primarily owing to increased light confinement (reduced radiative coupling). Instead, we observed (and predict, nonclassically) a near-constant broadening that is reduced slightly for very small gaps. This near-constant broadening results from an interplay (equation (4)) among the strong (classical) gap dependence of $\text{Re}(\tilde{\omega})$, the increase of the nonclassical perturbation strength (Fig. 3a) at smaller g values and the decrease of $\text{Im}(d_{\perp}^{\text{Au-AuO}_x})$ towards the infrared (Fig. 3c). Strikingly, the smallest gap does not produce the strongest nonclassical broadening (that is, $\text{Im}(\tilde{\omega}^{(1)})$), in contrast to the natural

expectation of monotonically increasing $\text{Im}(\tilde{\omega}^{(1)})$ with decreasing g . Instead, $\text{Im}(\tilde{\omega}^{(1)})$ is minimal there—a consequence of the near-vanishing magnitude of the strongly dispersive $\text{Im}(d_{\perp}^{\text{Au-AlO}_x})$ (Fig. 3c). This behaviour demonstrates the apparent breakdown of the empirical understanding of nonclassical broadening in nanostructures, known as Kreibig damping⁵, which holds that $\text{Im}(\tilde{\omega}^{(1)})$ is approximately proportional to $1/g$.

The observation of large nonclassical corrections in our coupled Au–Au setup frames the question: can nonclassical effects—which are often detrimental—be efficiently mitigated even in nanoscopic settings? To answer by example, we consider a hybrid dielectric–metal design, replacing the Au nanodisks with Si. Such hybrid configurations have been predicted to yield higher radiative efficiency with comparable overall plasmonic response²⁹ and have two key advantages for mitigating nonclassical effects: first, undoped Si is effectively a purely classical material (that is, $d_{\perp,||}^{\text{Si-AlO}_x} \simeq 0$) under the jellium approximation, as it lacks free electrons; second, high-index nanoresonators reduce the field intensity at the metal interface while maintaining confinement in the gap region. This hybridization can be exploited to reduce the nonclassical perturbation strength $\kappa_{\perp}^{\text{Au-AlO}_x}$ by an order of magnitude relative to that in the Au–Au design, as shown in Fig. 4a. Our measurements confirm this prediction: for Si nanodisks with $D \approx 104.4$ nm, we observe a high-quality scattering spectrum with a symmetric single-resonance feature for all gap sizes (Fig. 4b). The measured resonance frequencies (Fig. 4c) show only minor deviations from classical predictions, in both real and imaginary parts. Although the inclusion of nonclassical effects improves the experimental agreement, the overall shift remains small and comparable to the uncertainties owing to the intrinsic oxide thickness beneath the Si nanodisk (Supplementary Information section S12.C). By considering a range of nanodisk diameters (Fig. 4d), we reach an identical conclusion, even for the smallest considered gap (about 1.1 nm): classical scattering spectra agree well with measurements, and nonclassical corrections are minor relative to those in the Au–Au system. We found similar robustness across several additional gap sizes and diameters (Supplementary Information section S14).

Equation (2) suggests a complementary strategy for mitigating nonclassical effects: if the sign of $\text{Re}(d_{\perp}^{\tau})$ differs at distinct interfaces τ , the interface summation $\text{Re}(\sum_{\tau} \kappa_{\perp}^{\tau} d_{\perp}^{\tau})$ will partially cancel. Whereas noble metals are known to spill inwards ($\text{Re}(d_{\perp}) < 0$), simple metals (for example, Al) spill outwards ($\text{Re}(d_{\perp}) > 0$)¹⁹. We found experimental evidence for such a partial cancellation in a combined noble–simple-metal setup (Al nanodisks on a Au substrate; Supplementary Information section S15).

The mesoscopic framework presented here introduces a general approach for incorporating nonclassical effects in electromagnetic problems by a simple generalization of the associated boundary conditions. Our experiments show how to directly measure the nonclassical surface-response functions—the Feibelman d parameters—in general and technologically relevant plasmonic platforms. Our findings establish the Feibelman d parameters as first-class response functions of general utility. This calls for the compilation of databases of d parameters at interfaces of photonic and plasmonic prominence, analogous and complementary to the existing databases of local bulk permittivities.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1803-1>.

- Maxwell, J. C. A dynamical theory of the electromagnetic field. *Philos. Trans. R. Soc. Lond.* **155**, 459–512 (1865).
- Nielsen, M., Shi, X., Dichtl, P., Maier, S. A. & Oulton, R. F. Giant nonlinear response at a plasmonic nanofocus drives efficient four-wave mixing. *Science* **358**, 1179–1181 (2017).
- Chikkaraddy, R. et al. Single-molecule strong coupling at room temperature in plasmonic nanocavities. *Nature* **535**, 127–130 (2016).
- Akselrod, G. M. et al. Probing the mechanisms of large Purcell enhancement in plasmonic nanoantennas. *Nat. Photon.* **8**, 835–840 (2014).
- Kreibig, U. & Genzel, L. Optical absorption of small metallic particles. *Surf. Sci.* **156**, 678–700 (1985).
- Tiggesbäumker, J., Köller, L., Meiwes-Broer, K.-H. & Liebsch, A. Blue shift of the Mie plasma frequency in Ag clusters and particles. *Phys. Rev. A* **48**, R1749–R1752 (1993).
- Cottancin, E. et al. Optical properties of noble metal clusters as a function of the size: comparison between experiments and a semi-quantal theory. *Theor. Chem. Acc.* **116**, 514–523 (2006).
- Ciraci, C. et al. Probing the ultimate limits of plasmonic enhancement. *Science* **337**, 1072–1074 (2012).
- Schöll, J. A., Koh, A. L. & Dionne, J. A. Quantum plasmon resonances of individual metallic nanoparticles. *Nature* **483**, 421–427 (2012).
- Raza, S. et al. Blueshift of the surface plasmon resonance in silver nanoparticles studied with EELS. *Nanophotonics* **2**, 131–138 (2013).
- Boardman, A. D. *Electromagnetic Surface Modes* (John Wiley & Sons, 1982).
- Fernández-Domínguez, A. I., Wiener, A., García-Vidal, F. J., Maier, S. A. & Pendry, J. B. Transformation-optics description of nonlocal effects in plasmonic nanostructures. *Phys. Rev. Lett.* **108**, 106802 (2012).
- Raza, S., Bozhevolnyi, S. I., Wubs, M. & Mortensen, N. A. Nonlocal optical response in metallic nanostructures. *J. Phys. Condens. Matter* **27**, 183204 (2015).
- Zhu, W. et al. Quantum mechanical effects in plasmonic structures with subnanometre gaps. *Nat. Commun.* **7**, 11495 (2016).
- Skjølstrup, E. J. H., Søndergaard, T. & Pedersen, T. G. Quantum spill-out in few-nanometer metal gaps: effect on gap plasmons and reflectance from ultrasharp groove arrays. *Phys. Rev. B* **97**, 115429 (2018).
- Zuloaga, J., Prodan, E. & Nordlander, P. Quantum description of the plasmon resonances of a nanoparticle dimer. *Nano Lett.* **9**, 887–891 (2009).
- Teperik, T. V., Nordlander, P., Aizpurua, J. & Borisov, A. G. Robust subnanometric plasmon ruler by rescaling of the nonlocal optical response. *Phys. Rev. Lett.* **110**, 263901 (2013).
- Feibelman, P. J. Surface electromagnetic fields. *Prog. Surf. Sci.* **12**, 287–407 (1982).
- Liebsch, A. *Electronic Excitations at Metal Surfaces* (Springer, 1997).
- Halperin, W. P. Quantum size effects in metal particles. *Rev. Mod. Phys.* **58**, 533–606 (1986).
- Yan, W., Wubs, M. & Mortensen, N. A. Projected dipole model for quantum plasmonics. *Phys. Rev. Lett.* **115**, 137403 (2015).
- Christensen, T., Yan, W., Jauho, A.-P., Soljačić, M. & Mortensen, N. A. Quantum corrections in nanoplasmonics: shape, scale, and material. *Phys. Rev. Lett.* **118**, 157402 (2017).
- Apell, P. & Ljungbert, A. A general non-local theory for the electromagnetic response of a small metal particle. *Phys. Scr.* **26**, 113–118 (1982).
- Liebsch, A. Dynamical screening at simple-metal surfaces. *Phys. Rev. B* **36**, 7378–7388 (1987).
- Liebsch, A. Surface-plasmon dispersion and size dependence of Mie resonance: silver versus simple metals. *Phys. Rev. B* **48**, 11317–11328 (1993).
- Feibelman, P. J. Comment on “Surface plasmon dispersion of Ag”. *Phys. Rev. Lett.* **72**, 788 (1994).
- Lalanne, P., Yan, W., Vynck, K., Sauvan, C. & Hugonin, J.-P. Light interaction with photonic and plasmonic resonances. *Laser Photonics Rev.* **12**, 1700113 (2018).
- Yan, W., Faggiani, R. & Lalanne, P. Rigorous modal analysis of plasmonic nanoresonators. *Phys. Rev. B* **97**, 205422 (2018).
- Yang, Y., Miller, O. D., Christensen, T., Joannopoulos, J. D. & Soljačić, M. Low-loss plasmonic dielectric nanoresonators. *Nano Lett.* **17**, 3238–3245 (2017).
- Yang, J., Giessen, H. & Lalanne, P. Simple analytical expression for the peak-frequency shifts of plasmonic resonances for sensing. *Nano Lett.* **15**, 3439–3444 (2015).
- Jin, D. et al. Quantum-spillover-enhanced surface-plasmonic absorption at the interface of silver and high-index dielectrics. *Phys. Rev. Lett.* **115**, 193901 (2015).
- Groner, M. D., Elam, J. W., Fabreguette, F. H. & George, S. M. Electrical characterization of thin Al_2O_3 films grown by atomic layer deposition on silicon and various metal substrates. *Thin Solid Films* **413**, 186–197 (2002).
- Banerjee, A. et al. Optical properties of refractory metal based thin films. *Opt. Mater. Express* **8**, 2072–2088 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Dark-field microscopy

We built a table-top dark-field microscope (Fig. 2a) switchable between imaging and spectroscopy modes and with a 100–700× variable zoom, to measure $\tilde{\omega}$ from the optical response of the samples (Fig. 2b). Optical spectra were recorded at full zoom, capturing the scattered light from an ensemble of ≤ 100 nanodisks (ensemble—rather than single-particle—measurements are necessary owing to the associated low radiative efficiencies; see Supplementary Information section S10). Mutual coupling between nanodisks in the array is negligible, ensured by a lattice periodicity of 2 μm (in-plane filling factor $< 1\%$). This allows an isolated-particle treatment.

Structural characterization

The size distribution of the nanodisks was characterized systematically to adjust for the impact of inhomogeneous broadening in the measured scattering spectrum from the ensemble (Fig. 2c, Supplementary Information section S7). We measured the AlO_x gap size g using a variable-angle ultraviolet–visible ellipsometer and confirmed the results through cross-sectional transmission electron microscopy (see Fig. 2d and Supplementary Information section S8), finding good agreement with nominal ALD cycle expectations. A surface roughness of about 0.6 nm (root mean square) was measured for the Au substrate using atomic-force microscopy and was taken as the gap size uncertainty. Owing to the conformal nature of the ALD, such roughness should have negligible influence on the scattering spectra, as we verified by numerical simulations (Supplementary Information section S11). These detailed characterizations eliminate the main sources of geometric uncertainty in the mapping between calculated $\tilde{\omega}^{(0)}$ and measured $\tilde{\omega}$, which is a necessity for an accurate evaluation of the nonclassical shift $\tilde{\omega} - \tilde{\omega}^{(0)}$.

Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgements We thank J. Daley, S. E. Kooi and M. Mondol for assistance in sample fabrication and measurement. We thank F. Niroui and T. Zhu for lending us equipment. We acknowledge discussions with V. Bulović, O. D. Miller and N. A. Mortensen. We thank P. Rebusco for reading and editing the manuscript. K.K.B. thanks A. Chu and J. Wanapun for support. This work was partly supported by the US Army Research Office through the Institute for Soldier Nanotechnologies under contract number W911NF-18-2-0048 and W911NF-13-D-0001, and by Air Force Office of Scientific Research (AFOSR) under grant contract number FA9550-18-1-0436. Y.Y. was partly supported by the MRSEC Program of the National Science Foundation under grant number DMR-1419807. D.Z. was supported by a National Science Scholarship from A*STAR, Singapore. W.Y. was supported by Programme IdEx Bordeaux-LAPHIA (grant number ANR-10-IDEX-03-02) and project ‘Resonance’ (grant number ANR-16-CE24-0013) of the French National Agency for Research (ANR). M.Z. was supported by the National Natural Science Foundation of China (grant number 11574078) and the China Scholarship Council. T.C. was supported by the Danish Council for Independent Research (grant number DFFC6108-00667).

Author contributions Y.Y. and T.C. conceived the idea. D.Z. fabricated the samples. Y.Y. and D.Z. designed the experiment, built the setup, conducted the scattering measurements and performed the ellipsometry. T.C. derived the mesoscopic boundary conditions. Y.Y., W.Y. and T.C. developed the numerical methods and Y.Y. performed the numerical calculations. W.Y. proposed the auxiliary-potential method, performed density functional theory calculations and implemented the QNM-based perturbation analysis. D.Z. performed the atomic-force microscopy measurement. A.A. and D.Z. performed the transmission electron microscopy measurement. D.Z. and M.Z. characterized nanoparticle size statistics. Y.Y., D.Z., W.Y. and T.C. analysed the data. Y.Y., D.Z. and T.C. drafted the manuscript with extensive input from all authors. J.D.J., P.L., T.C., K.K.B. and M.S. supervised the project.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1803-1>.

Correspondence and requests for materials should be addressed to Y.Y., D.Z. or T.C.

Peer review information *Nature* thanks Michal Lipson and Javier Aizpurua for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Data-driven design of metal–organic frameworks for wet flue gas CO₂ capture

<https://doi.org/10.1038/s41586-019-1798-7>

Received: 24 August 2018

Accepted: 18 September 2019

Published online: 11 December 2019

Peter G. Boyd^{1,10}, Arunraj Chidambaram^{1,10}, Enrique García-Díez^{2,10}, Christopher P. Ireland¹, Thomas D. Daff^{3,8}, Richard Bounds⁴, Andrzej Gładysiak¹, Pascal Schouwink⁵, Seyed Mohamad Moosavi¹, M. Mercedes Maroto-Valer², Jeffrey A. Reimer^{4,7}, Jorge A. R. Navarro⁶, Tom K. Woo^{3*}, Susana Garcia^{2*}, Kyriakos C. Stylianou^{1,9*} & Berend Smit^{1*}

Limiting the increase of CO₂ in the atmosphere is one of the largest challenges of our generation¹. Because carbon capture and storage is one of the few viable technologies that can mitigate current CO₂ emissions², much effort is focused on developing solid adsorbents that can efficiently capture CO₂ from flue gases emitted from anthropogenic sources³. One class of materials that has attracted considerable interest in this context is metal–organic frameworks (MOFs), in which the careful combination of organic ligands with metal-ion nodes can, in principle, give rise to innumerable structurally and chemically distinct nanoporous MOFs. However, many MOFs that are optimized for the separation of CO₂ from nitrogen^{4–7} do not perform well when using realistic flue gas that contains water, because water competes with CO₂ for the same adsorption sites and thereby causes the materials to lose their selectivity. Although flue gases can be dried, this renders the capture process prohibitively expensive^{8,9}. Here we show that data mining of a computational screening library of over 300,000 MOFs can identify different classes of strong CO₂-binding sites—which we term ‘adsorbaphores’—that endow MOFs with CO₂/N₂ selectivity that persists in wet flue gases. We subsequently synthesized two water-stable MOFs containing the most hydrophobic adsorbaphore, and found that their carbon-capture performance is not affected by water and outperforms that of some commercial materials. Testing the performance of these MOFs in an industrial setting and consideration of the full capture process—including the targeted CO₂ sink, such as geological storage or serving as a carbon source for the chemical industry—will be necessary to identify the optimal separation material.

Different strategies have been developed to mitigate the negative effects of water on the CO₂/N₂ separation selectivity in MOF materials. For example, some MOFs have open metal sites at which amines can be attached, taking advantage of the specific amine chemistry that is also used in conventional amine scrubbing^{10–13}. A previous screening study¹⁴ investigated whether MOFs could adsorb CO₂ in the presence of water, and the results suggested that such MOFs could be de novo designed. In this work, we develop a systematic strategy for the design and preparation of custom-made MOFs that can capture carbon from wet flue gases. Our design methodology is inspired by the rational design of drug molecules, in which organic molecules that fit well into the binding pocket of a protein are mined from databases of known chemicals^{15,16}. The difference in our case is that the ‘drug molecule’ is known (that is, CO₂), but the substrate that binds it optimally (the MOF)

is not. We therefore generated a library of 325,000 hypothetical MOFs, and screened each material for its CO₂/N₂ selectivity and its CO₂ working capacity. The chemical building blocks used in the generation of these materials are shown in Extended Data Figs. 1 and 2. Figure 1a shows that 8,325 hypothetical materials possess a working capacity for CO₂ greater than 2 mmol g^{−1} and a CO₂/N₂ selectivity greater than 50—performance that surpasses that of zeolite 13X under dry conditions¹⁷.

A key part of drug design is to analyse the optimally binding molecules for a common feature or spatial arrangement of atoms at the binding site, which is known as the pharmacophore¹⁵. In analogy with this, we coin the term ‘adsorbaphore’ to describe the common pore shape and chemistry of a binding site in a MOF that provides optimal interactions to preferentially bind to a particular guest molecule, in this case CO₂. From our top-ranked 8,325 materials, we identified

¹Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, Valais (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland. ²Research Centre for Carbon Solutions (RCCS), School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK. ³Department of Chemistry and Biomolecular Science, University of Ottawa, Ottawa, Ontario, Canada. ⁴Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, CA, USA. ⁵Institut des Sciences et Ingénierie Chimiques (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. ⁶Departamento de Química Inorgánica, Universidad de Granada, Granada, Spain. ⁷Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁸Present address: Department of Engineering, University of Cambridge, Cambridge, UK. ⁹Present address: Department of Chemistry, Oregon State University, Corvallis, OR, US. ¹⁰These authors contributed equally: Peter G. Boyd, Arunraj Chidambaram, Enrique García-Díez. *e-mail: twoo@uottawa.ca; S.Garcia@hw.ac.uk; kyriakos.stylianou@oregonstate.edu; berend.smit@epfl.ch

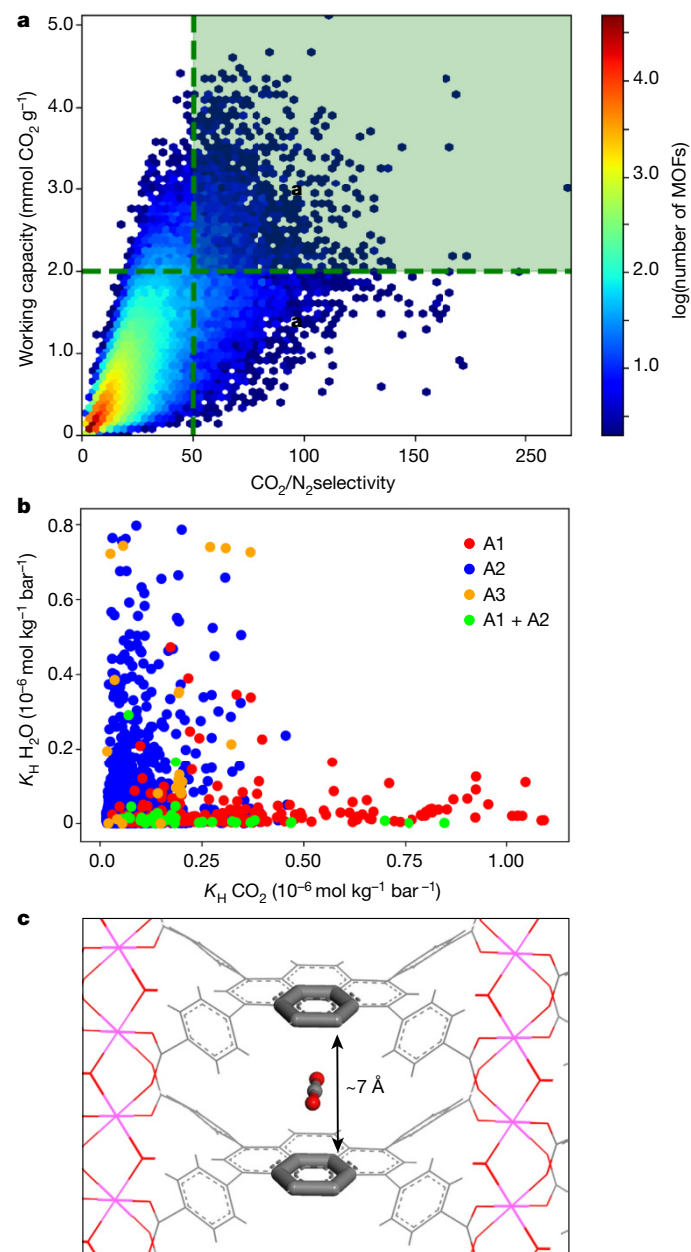


Fig. 1 | Computational screening of MOFs for strong CO₂ adsorption and selectivity. **a**, Results from the screening of 325,000 hypothetical MOFs under conditions that mimic post-combustion capture (adsorption at ambient temperature and 1 bar with a molar ratio of 15:85 CO₂/N₂ and regeneration at 363 K and 0.1 bar). The materials in the green box were selected for more refined screening and adsorbaphore identification; the colour-coding represents the number of MOFs according to the colour bar on the right. **b**, The H₂O affinity of the top-performing materials, as characterized by a plot of the Henry coefficients (K_H) for H₂O against those of CO₂. The colours represent the three different adsorbaphores found in the top-performing structures: A1, parallel aromatic rings; A2, metal–oxygen bridges; A3, open metal sites. Some materials have both A1 and A2 sites. **c**, The adsorbaphore containing parallel aromatic rings (A1), which was discovered using the feature-recognition algorithm.

106,680 such CO₂-binding sites (see Extended Data Fig. 3 for examples). A similarity analysis of these binding sites revealed three main classes of adsorbaphore: A1, two parallel aromatic rings with interatomic spacings of approximately 7 Å (31% of all binding sites); A2, metal–oxygen–metal bridges (32%); and A3, open metal sites (21%) (see Supplementary Information for details). Subsequently, we screened the materials that possessed these adsorbaphores for their affinity for water. Figure 1b

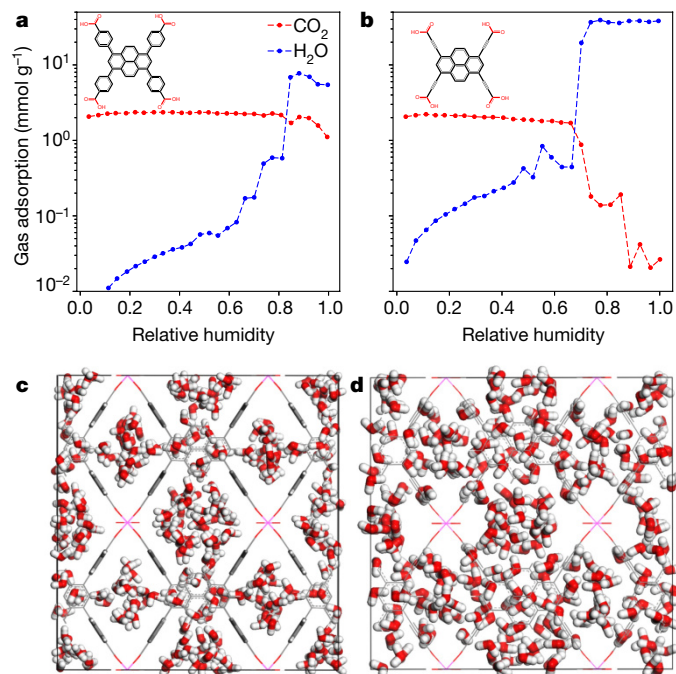


Fig. 2 | The effects of water on different MOFs with the same adsorbaphore.

a, b, Simulated adsorption of a ternary mixture of CO₂/N₂/H₂O at 313 K by hypothetical MOFs **m8o67** (**a**) and **m8o71** (**b**). The partial pressure of CO₂ was held at 0.15 bar as the relative humidity was increased. The N₂ uptake was negligible and is not shown. **c, d**, Visualization of the water loading at 100% relative humidity in **m8o67** (**c**) and **m8o71** (**d**). The benzoate groups are represented by grey sticks; water is shown using red and white space-filling atoms. In **m8o67**, the benzoate groups perpendicular to the plane of the figure prevent hydrogen-bond formation across the adsorbaphores.

shows the Henry coefficient for water in these high-performing materials. Analysis of the data shows that the materials with the parallel aromatic rings (A1) have a low Henry coefficient for H₂O, whereas those with metal–oxygen bridges (A2) and open metal sites (A3) tend to have higher Henry coefficients (Fig. 1b). A graphical representation of the different adsorbaphores is presented in Extended Data Fig. 4. Comparison of the binding energies—computed using density functional theory—for the adsorbaphore shown in Fig. 1c indicates a preference for CO₂ (−10.2 kcal mol^{−1}) over N₂ and H₂O by 2.7 and 1.5 kcal mol^{−1}, respectively (see Extended Data Table 1). The parallel aromatic rings provide a near-optimum interaction with all three atoms of CO₂, whereas for H₂O the lack of hydrogen-bonding sites limits its binding energy.

The next step was to identify a subclass of MOFs in our library that contains the preferred adsorbaphore. From an experimental point of view, MOFs with the frz topology—characterized by tetra-carboxylated organic ligands coordinated to one-dimensional metal–oxygen rods—are an attractive starting point. One such example has been synthesized with indium as a metal node, resulting in a structurally stable, non-breathing MOF¹⁸. In this topology, the metal rods provide an ideal scaffolding to which we can attach our adsorbaphore. By varying the metal ion we have some flexibility to tune the distance between the aromatic rings. Our calculations predict that the ideal adsorbaphore distance of 6.5–7.0 Å—which was determined by adjusting the spacing of the aromatic rings incrementally (Extended Data Fig. 5)—can be approached if In(III) is replaced by Al(III) (Extended Data Table 2). In addition, aluminium is an attractive choice because it is an abundant metal and it ensures a strong bond with the carboxylate O-atoms of the ligands¹⁹; this considerably improves the thermal and hydrolytic stability of a MOF^{20,21}.

We generated a library of 35 isorecticular materials using our MOF-generation algorithm²², and from the mixture isotherms we computed the

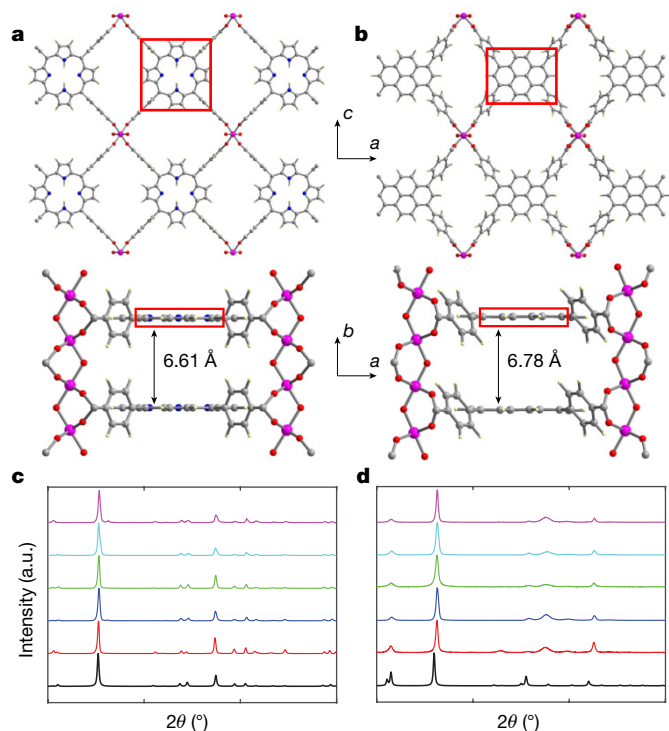


Fig. 3 | Structural representation and stability of [Al-PMOF] and [Al-PyrMOF]. **a, b,** Ball-and-stick representation of the structures of [Al-PMOF] (**a**) and [Al-PyrMOF] (**b**). The orientation of the tetracarboxylate ligands around the Al(III) rods results in the generation of the three-dimensional non-interpenetrated structures containing the adsorbaphore (red box). Atom colour code: pink, Al; grey, C; blue, N; red, O; pale yellow, H. **c, d,** Laboratory powder X-ray diffraction patterns of [Al-PMOF] (**c**) and [Al-PyrMOF] (**d**). Black, simulated; red, as-synthesized material; blue, acetone-exchanged material; green, activated material; sky blue, activated material immersed in liquid water for 7 days; pink, activated material exposed to a controlled atmosphere of nitric acid vapour for 3 h.

CO₂/N₂ selectivity of the materials in dry and wet flue gases (Extended Data Figs. 6 and 7). Our calculations show that all of our predicted materials maintain an excellent selectivity at low pressures, and in about 75% of these materials the selectivity was not influenced by the presence of water under flue-gas conditions. The concept of an adsorbaphore focuses on the design of an adsorption site that optimizes selectivities at low pressure. At higher partial pressures of water, its adsorption is dominated by the energetics of hydrogen-bond formation. Further analyses showed that, for the materials that maintain a high CO₂ uptake at high humidity, it is the pore shape that frustrates the formation of these hydrogen bonds. This is illustrated in Fig. 2a, b, which compares the effect of water on the CO₂ uptake of two materials that have the same adsorbaphore but different pore structures (hypothetical MOFs **m8o67** and **m8o71**). Figure 2a shows that **m8o67** is resistant to H₂O flooding: even at a relative humidity of approximately 85%, we find that H₂O has only a small effect on CO₂ capacity. Conversely, **m8o71** completely loses its CO₂ capacity at 60% relative humidity (Fig. 2b). In Fig. 2c, d we visualize the hydrogen-bond network that is formed at 100% relative humidity in both materials. For **m8o71** we see a complete hydrogen-bonding network (Fig. 2d), whereas for **m8o67** (Fig. 2c) we observe a less extensive network; the benzoate groups that separate the adsorbaphores frustrate the formation of a complete hydrogen-bonding network.

On the basis of these predictions, we synthesized two frz-based MOFs using organic ligands that possess the water-frustrating properties reported above: **Al-PMOF**¹⁹ (**m8o66**) and **Al-PyrMOF** (**m8o67**). These MOFs are based on one-dimensional rods of Al(III) linked by

TCPP (tetrakis(4-carboxyphenyl)porphyrin) and TBAPy (1,3,6,8-tetrakis(*p*-benzoic acid)pyrene) ligands, respectively (Fig. 3a, b). Figure 3c, d shows no loss of crystallinity upon activation as well as upon exposure to different harsh conditions, including immersion in water for 7 days.

By identifying adsorbaphores in these hypothetical materials, we assume that our *in silico* screening method can correctly predict the structure of a MOF, its adsorption properties, and the nature of the binding sites of CO₂ and H₂O. We are able to test these assumptions for **Al-PMOF** and **Al-PyrMOF**. In Fig. 4a we show that the experimental and predicted CO₂ and N₂ adsorption isotherms are in good agreement. The CO₂ binding positions in the adsorbaphore and the effect of H₂O are more challenging to observe experimentally. The siting of CO₂ was studied using *in situ* CO₂-loading powder X-ray diffraction. Upon loading, we observed a considerable change in the intensity and peak position of the Bragg reflections (see Supplementary Fig. 2.1). Subsequent Rietveld refinement and Fourier analysis²³ revealed the preferred locations of CO₂ in the pores of **Al-PMOF**, as shown in Fig. 4b. These results confirm that CO₂ preferentially adsorbs in the adsorbaphore.

The effect of water on the siting of CO₂ has been further addressed by solid-state nuclear magnetic resonance (NMR) analysis. Under magic-angle spinning (MAS), high-resolution ¹³C NMR chemical shifts are very sensitive to changes in the chemical environment. The ¹³C NMR spectra of **Al-PyrMOF** and **Al-PMOF** are provided in Extended Data Fig. 8, which also shows the assignment of the peaks to specific atoms on the MOF. The chemical shifts associated with the atoms of the adsorbaphore (inset) are shown in Fig. 4c as a function of the water concentration. At low levels the adsorbaphore atoms experience no change in chemical environment with water loading, whereas at the highest water loadings there are modest changes in the carbon-13 chemical shifts of only those atoms that are close to the aluminium-coordinated carboxylate groups next to the adsorbaphore (carbons B and F in Fig. 4c). This broadening is consistent with dipolar broadening from proximate water molecules, thus confirming that the adsorbaphore itself is not a preferential adsorption site for H₂O.

Our simulations predict that CO₂ adsorbed in the adsorbaphore is insulated from the adsorption of water. Because the ¹³C NMR spectrum of adsorbed ¹³CO₂ is extremely sensitive to the proximity of water molecules in terms of chemical shift and line broadening, any disruption of the chemical environment of adsorbed CO₂ by water should be immediately apparent. Figure 4d shows that the chemical shift of the adsorbed ¹³CO₂ is independent of water content, although a broadening of the ¹³C NMR peak is observed with increasing humidity. If this broadening is due to the proximity of the protons in water, it should disappear if the experiments are repeated using D₂O; however, it does not (Fig. 4d). This observation corroborates our simulation results (Fig. 2c), confirming that water has only a limited effect on CO₂ adsorption in **Al-PyrMOF**.

The ability of these materials to capture CO₂ from wet flue gases is of important practical concern. We therefore used a breakthrough experiment to determine the capture capacity of both **Al-PMOF** and **Al-PyrMOF** for a mixture of CO₂/N₂ under dry- and humid-conditions²⁴ (Fig. 4e). These results confirm the predictions of the simulations (Extended Data Fig. 7): humidity in the flue gases has only a minimal influence on the capture capacity of **Al-PMOF**, whereas for **Al-PyrMOF** the capture capacity is in fact enhanced. Furthermore, repeated cycling²⁵ (Fig. 4f) does not result in degradation of the material or a change in separation performance. It is instructive to compare the performance of our materials with that of a set of reference materials, including those that are commercially available—such as zeolite 13X and activated carbon—and a water stable, amino-functionalized MOF, UiO-66-NH₂. The capture capacity of these reference materials lies between that of **Al-PyrMOF** and **Al-PMOF** in dry flue gases; however, unlike our MOFs, their performance reduces considerably in humid flue gases. Although our materials do not have the highest reported

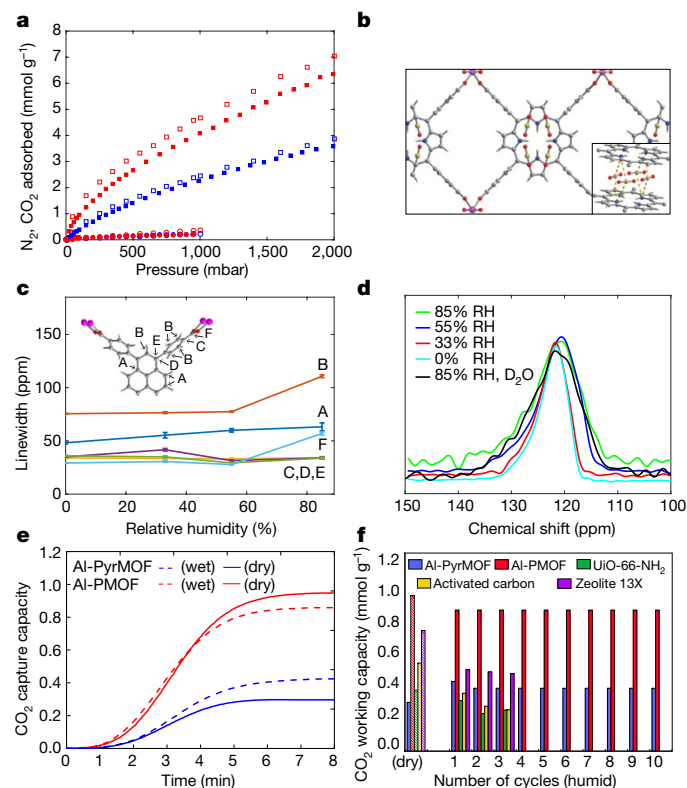


Fig. 4 | CO₂ adsorption, ¹³C cross-polarization MAS NMR and breakthrough experiments for [Al-PMOF] and [Al-PyrMOF]. **a**, Experimental (filled) and computational (open) single-component adsorption isotherms for CO₂ (squares) and N₂ (circles) adsorption collected on activated [Al-PMOF] (red) and [Al-PyrMOF] (blue) at 313 K. **b**, Rietveld refinement of the X-ray diffraction data (Supplementary Information) revealed that CO₂ binding in [Al-PMOF] occurs between the porphyrin cores—that is, in the adsorbaphore. **c**, The linewidth of each carbon peak of the TBAPy ligand in [Al-PyrMOF] in the ¹³C cross-polarization MAS spectrum, plotted as a function of relative humidity. Each carbon atom in the ligand is labelled in the inset. **d**, Linewidths extracted from the ¹³C static NMR spectra of ¹³CO₂ loaded in [Al-PyrMOF], plotted against relative humidity (RH). **e**, CO₂ capture capacity profiles for [Al-PyrMOF] and [Al-PMOF] during breakthrough experiments under dry and humid (85% relative humidity) conditions, with 85/15 v/v of N₂/CO₂ (313 K and 1 bar). **f**, Benchmarking the CO₂ working capacity of [Al-PyrMOF] and [Al-PMOF] against UiO-66-NH₂, activated carbon and zeolite 13X under dry and humid (85% relative humidity) conditions, with 85/15 v/v of N₂/CO₂ (313 K and 1 bar). For wet flue gases, we studied the performance stability after 3 cycles for reference materials, and after 10 cycles for [Al-PyrMOF] and [Al-PMOF].

working capacities¹⁴, it is encouraging to see that, in wet flue gases, **Al-PMOF** outperforms commercial materials such as zeolite 13X and activated carbon.

Large-scale screening of databases of hypothetical MOFs for various gas separation and storage applications has been reported previously^{26–29}; however, here we have focused on identifying binding pockets—or structural motifs termed adsorbaphores—as synthetic targets, rather than whole materials. This enhances the synthetic viability of the approach, as demonstrated by the identification of one new material with the targeted adsorbaphore that was synthesized and shown to adsorb CO₂ as predicted. The concept of linking computational screening with the synthesis of the corresponding materials through such adsorbaphores should be applicable to other gas separations of increasing complexity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at <https://doi.org/10.1038/s41586-019-1798-7>.

- Smit, B., Reimer, J. R., Oldenburg, C. M. & Bourg, I. C. *Introduction to Carbon Capture and Sequestration* (Imperial College Press, 2014).
- Bui, M. et al. Carbon capture and storage (CCS): the way forward. *Energy Environ. Sci.* **11**, 1062–1176 (2018).
- D'Alessandro, D. M., Smit, B. & Long, J. R. Carbon dioxide capture: prospects for new materials. *Angew. Chem. Int. Ed.* **49**, 6058–6082 (2010).
- Sumida, K. et al. Carbon dioxide capture in metal–organic frameworks. *Chem. Rev.* **112**, 724–781 (2012).
- Furukawa, H., Cordova, K. E., O’Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal–organic frameworks. *Science* **341**, 1230–1234 (2013).
- Huck, J. M. et al. Evaluating different classes of porous materials for carbon capture. *Energy Environ. Sci.* **7**, 4132–4146 (2014).
- Mason, J. A., Sumida, K., Herm, Z. R., Krishna, R. & Long, J. R. Evaluating metal–organic frameworks for post-combustion carbon dioxide capture via temperature swing adsorption. *Energy Environ. Sci.* **4**, 3030–3040 (2011).
- Li, G. et al. Capture of CO₂ from high humidity flue gas by vacuum swing adsorption with zeolite 13X. *Adsorption* **14**, 415–422 (2008).
- Merel, J., Clause, M. & Meunier, F. Experimental investigation on CO₂ post-combustion capture by indirect thermal swing adsorption using 13X and 5A zeolites. *Ind. Eng. Chem. Res.* **47**, 209–215 (2008).
- Milner, P. J. et al. A diamine-appended metal–organic framework enabling efficient CO₂ capture from coal flue gas via a mixed adsorption mechanism. *J. Am. Chem. Soc.* **139**, 13541–13553 (2017).
- McDonald, T. M. et al. Cooperative insertion of CO₂ in diamine-appended metal–organic frameworks. *Nature* **519**, 303–308 (2015).
- Flaig, R. W. et al. The chemistry of CO₂ capture in an amine-functionalized metal–organic framework under dry and humid conditions. *J. Am. Chem. Soc.* **139**, 12125–12128 (2017).
- Couck, S. et al. An amine-functionalized MIL-53 metal–organic framework with large separation power for CO₂ and CH₄. *J. Am. Chem. Soc.* **131**, 6326–6327 (2009).
- Chanut, N. et al. Screening the effect of water vapour on gas adsorption performance: application to CO₂ capture from flue gas in metal–organic frameworks. *ChemSusChem* **10**, 1543–1553 (2017).
- Wolber, G., Seidel, T., Bendix, F. & Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* **13**, 23–29 (2008).
- Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W., Jr. Computational methods in drug discovery. *Pharmacol. Rev.* **66**, 334–395 (2013).
- Ho, M. T., Allinson, G. W. & Wiley, D. E. Reducing the cost of CO₂ capture from flue gases using pressure swing adsorption. *Ind. Eng. Chem. Res.* **47**, 4883–4890 (2008).
- Stylianos, K. C. et al. A guest-responsive fluorescent 3D microporous metal–organic framework derived from a long-lifetime pyrene core. *J. Am. Chem. Soc.* **132**, 4119–4130 (2010).
- Fateeva, A. et al. A water-stable porphyrin-based metal–organic framework active for visible-light photocatalysis. *Angew. Chem. Int. Ed.* **51**, 7440–7444 (2012).
- Loiseau, T. et al. A rationale for the large breathing of the porous aluminum terephthalate (MIL-53) upon hydration. *Chem. Eur. J.* **10**, 1373–1382 (2004).
- Reinsch, H. & Stock, N. High-throughput studies of highly porous Al-based MOFs. *Microporous Mesoporous Mater.* **171**, 156–165 (2013).
- Boyd, P. G. & Woo, T. K. A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm* **16**, 3777–3792 (2016).
- Carrington, E. J., Vitorica-Yrezabal, I. J. & Brammer, L. Crystallographic studies of gas sorption in metal–organic frameworks. *Acta Crystallogr. B* **70**, 404–422 (2014).
- García, S. et al. Breakthrough adsorption study of a commercial activated carbon for pre-combustion CO₂ capture. *Chem. Eng. J.* **171**, 549–556 (2011).
- García, S., Gil, M. V., Pis, J. J., Rubiera, F. & Pevida, C. Cyclic operation of a fixed-bed pressure and temperature swing process for CO₂ capture: experimental and statistical analysis. *Int. J. Greenhouse Gas Control* **12**, 35–43 (2013).
- Lin, L.-C. et al. In silico screening of carbon-capture materials. *Nat. Mater.* **11**, 633–641 (2012).
- Wilmer, C. E. et al. Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.* **4**, 83–89 (2012).
- Boyd, P. G., Lee, Y. & Smit, B. Computational development of the nanoporous materials genome. *Nat. Mater. Rev.* **2**, 17037 (2017).
- Yazaydin, A. O. et al. Screening of metal–organic frameworks for carbon dioxide capture from flue gas using a combined experimental and modeling approach. *J. Am. Chem. Soc.* **131**, 18198–18199 (2009).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Data availability

The computed data and hypothetical materials that were used in this Article are provided free of charge on the Materials Cloud (<https://doi.org/10.24435/materialscloud:2018.0016/v3>). An interactive version of Fig. 1a, b can also be found on this site. Data that are not included in the paper are available upon reasonable request to the corresponding authors.

Code availability

Topology Based Crystal Constructor (ToBasCCo), the Python program used to build hypothetical MOFs, is hosted on GitHub at <https://github.com/peteboyd/tobascco>. The Python code that compares common chemical features between fragments is also provided on GitHub at <https://github.com/peteboyd/adsorbaphore> and is dependent on a C library called MCQD which performs the maximum clique detection of the chemical graphs. An interface between Python and C for this is provided here at https://github.com/peteboyd/mcqd_api. The Automatic Binding Site Locator (ABSL) program, used to identify CO₂ binding sites in each MOF, is part of a broader Python-based code used to facilitate simulations of porous materials called Fully Automated Adsorption Analysis in Porous Solids (FA³PS). This is available on Bit-Bucket at <https://bitbucket.org/tdaff/automation>.

Acknowledgements K.C.S. was supported by the Swiss National Science Foundation (SNF) under the Ambizione Energy Grant n.PZENP2_166888, P.G.B. and B.S. by the European Research Council (ERC) Advanced Grant (grant agreement no. 666983, MaGic) and the National Center of Competence in Research (NCCR), Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL). A.C., C.P.I., and S.M.M. were supported by European Union's Horizon 2020 research and innovation programme under grant agreement no. 760899 (GENESIS). R.B. and J.A.R. were supported by the Center for Gas

Separations Relevant to Clean Energy Technologies, an Energy Frontier Research Center funded by the Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences under award DE-SC0001015. E.G.-D. is supported by the European Commission under the Research Fund for Coal and Steel (RFCS) Programme (project no. 709741). M.M.M.-V. and S.G. acknowledge the financial support from the Engineering and Physical Sciences Research Council (EP/N024540/1) and the Research Centre for Carbon Solutions (RCCS) at Heriot-Watt University. The authors thank the Swiss Norwegian Beamlines of ESRF for beamtime allocation at BM31 for the in-situ CO₂ loading and variable temperature powder X-ray diffraction experiments. This work was supported by a grant from the Swiss National Supercomputing Center (CSCS) under project no. s761, as well as resources from the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. J.A.R.N. acknowledges Spanish MINECO (CTQ2017-84692-R) and EU Feder funding. P.G.B., T.D.D. and T.K.W. would like to thank NSERC of Canada for financial support and Compute Canada for computing resources. S.G., J.A.R., and B.S. acknowledge support during the final stage of the work of the ACT-PrISMa project, which has received joint funding from BEIS, NERC and EPSRC (UK), funding from the Division of CCS R&D, US Department of Energy (USA), and funding from the Office Fédéral de l'Energie (Switzerland).

Author contributions P.G.B. and T.K.W. developed the adsorbaphore similarity algorithm and developed the library of hypothetical MOFs. P.G.B., T.K.W. and B.S. characterised the adsorbaphores; T.D.D. carried out the initial computational screening and developed the MOF-CO₂ binding site identification algorithm. S.M.M. carried out the pore similarity analysis; A.C., C.P.I. and K.C.S. synthesized and characterized the materials. The breakthrough experiments were carried out by E.G.-D., C.P.I., A.C., M.M.M.-V., J.A.R.N. and S.G. The NMR experiments were carried out by R.B. and J.A.R.; X-ray analysis was carried out by A.G. and P.S. All authors contributed to the analysis of the data and the writing of the manuscript.

Competing interests K.C.S., B.S., A.C., P.G.B. and T.K.W. have filed an international patent application (no. 18 168 544.7) that relates to water-stable polyaromatic MOF materials for CO₂ separation from flue gas and natural gas streams.

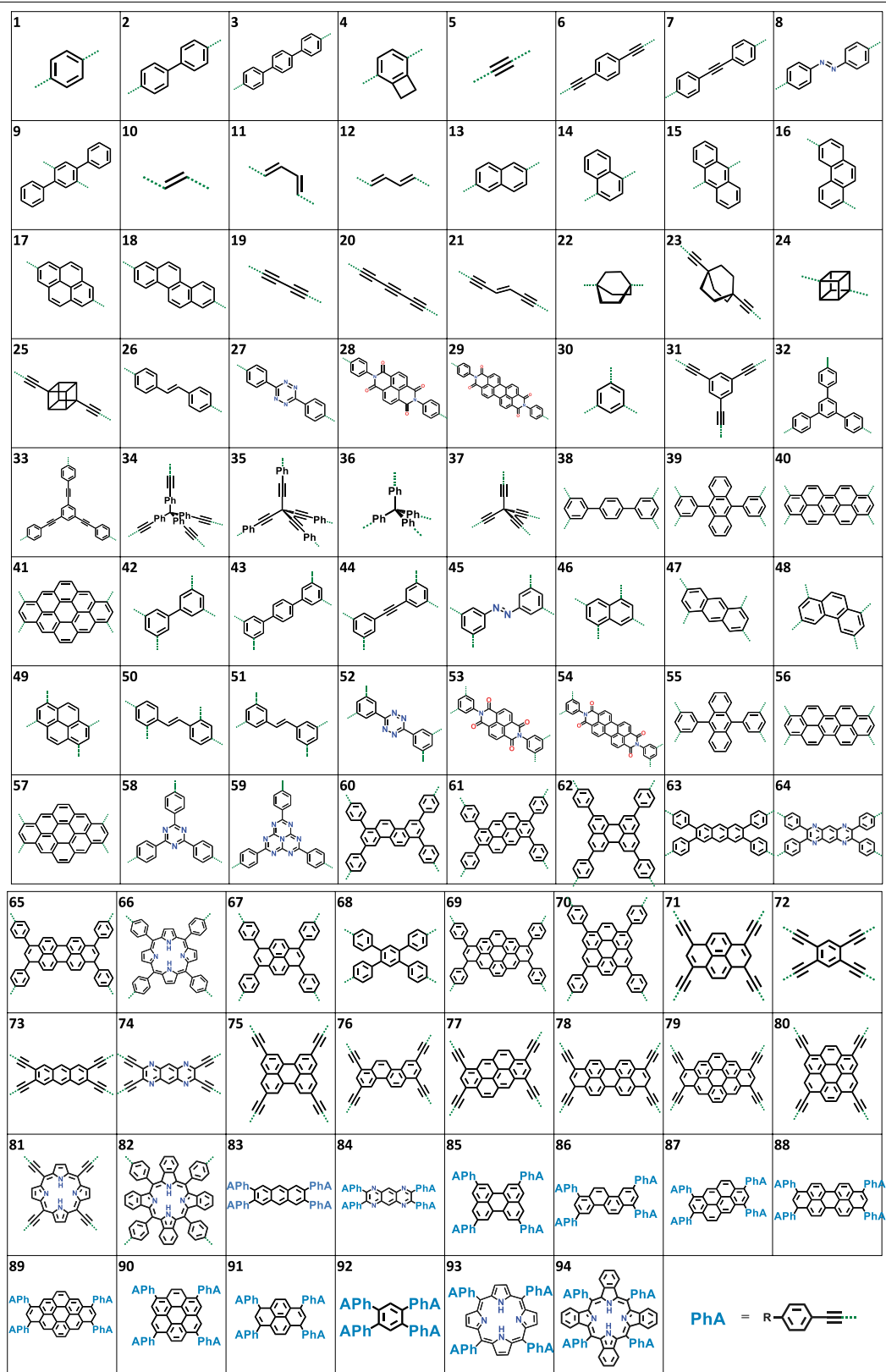
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1798-7>.

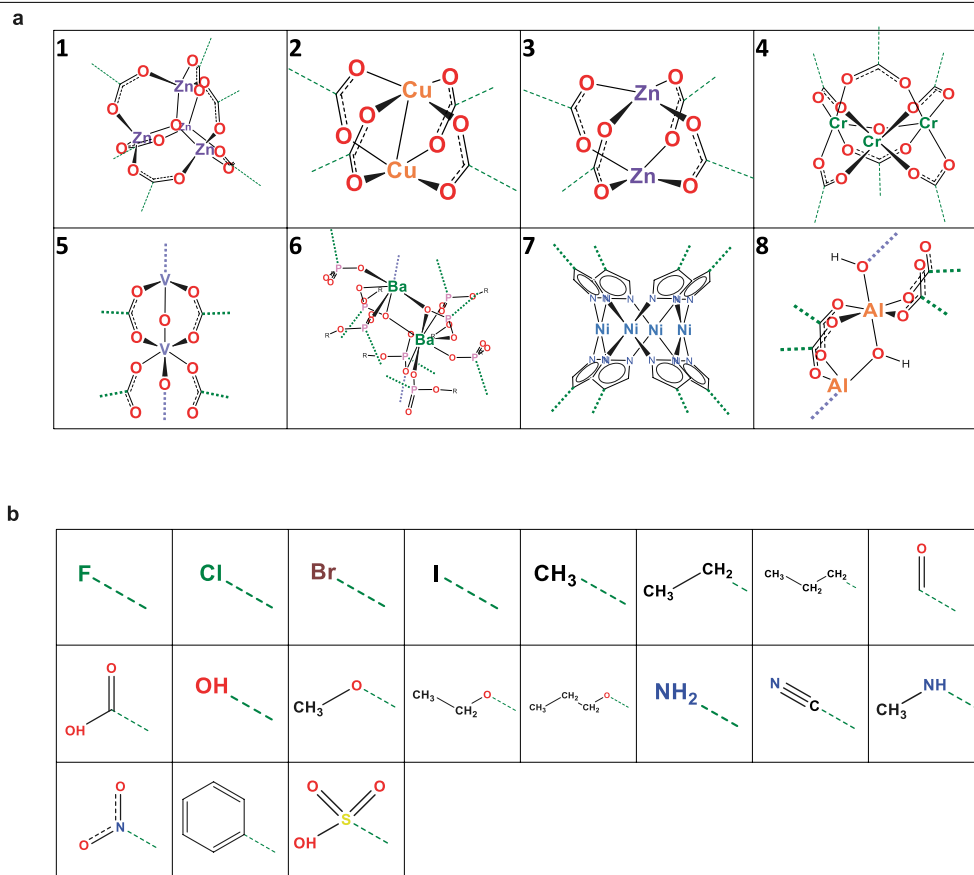
Correspondence and requests for materials should be addressed to T.K.W., S.G., K.C.S. or B.S.

Peer review information *Nature* thanks Philip Llewellyn and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

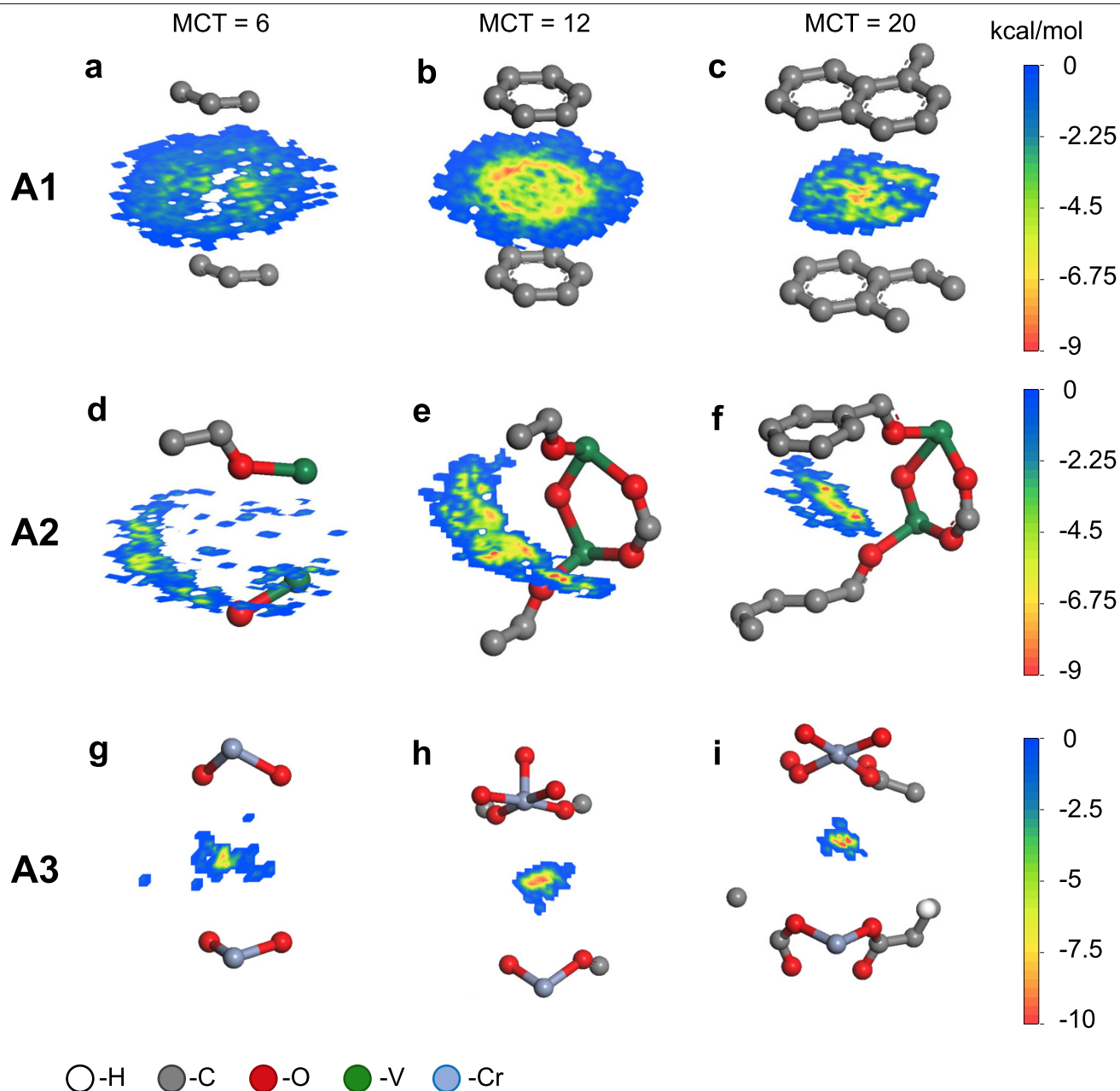


Extended Data Fig. 1 | Hypothetical material generation (1). The organic secondary building units (SBUs) used in the generation of the hypothetical MOF database.



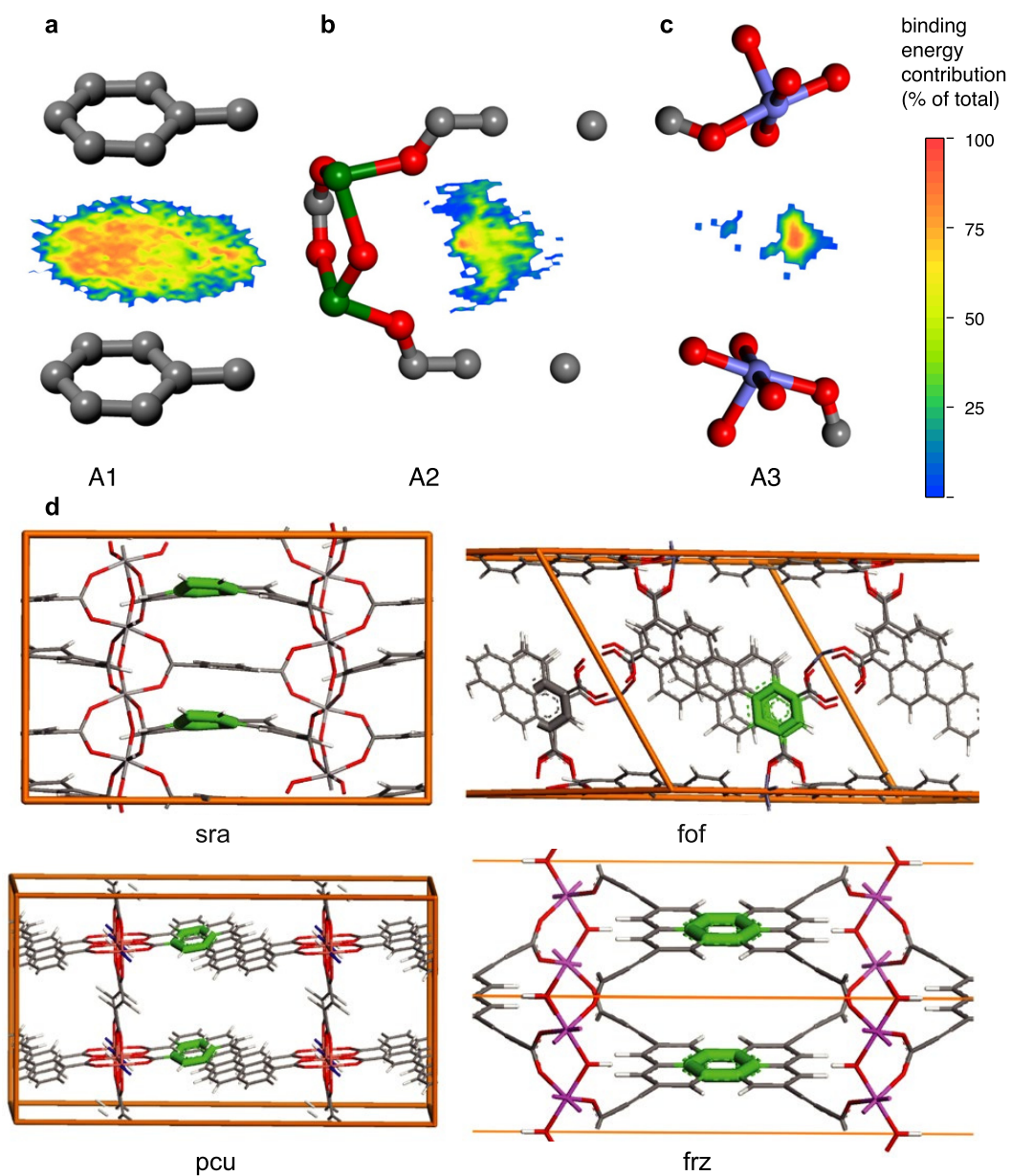
Extended Data Fig. 2 | Hypothetical material generation (2). **a**, Metal SBUs used in the generation of the hypothetical MOF database. **b**, Functional groups used to decorate the unfunctionalized hypothetical MOFs in the database. We denote a hypothetical material as **mXoYY**, where X refers to the metal SBU

shown in **a** and YY refers to the organic SBU shown in Extended Data Fig. 1. Functional groups were decorated onto the base hypothetical materials using an internal numbering system.



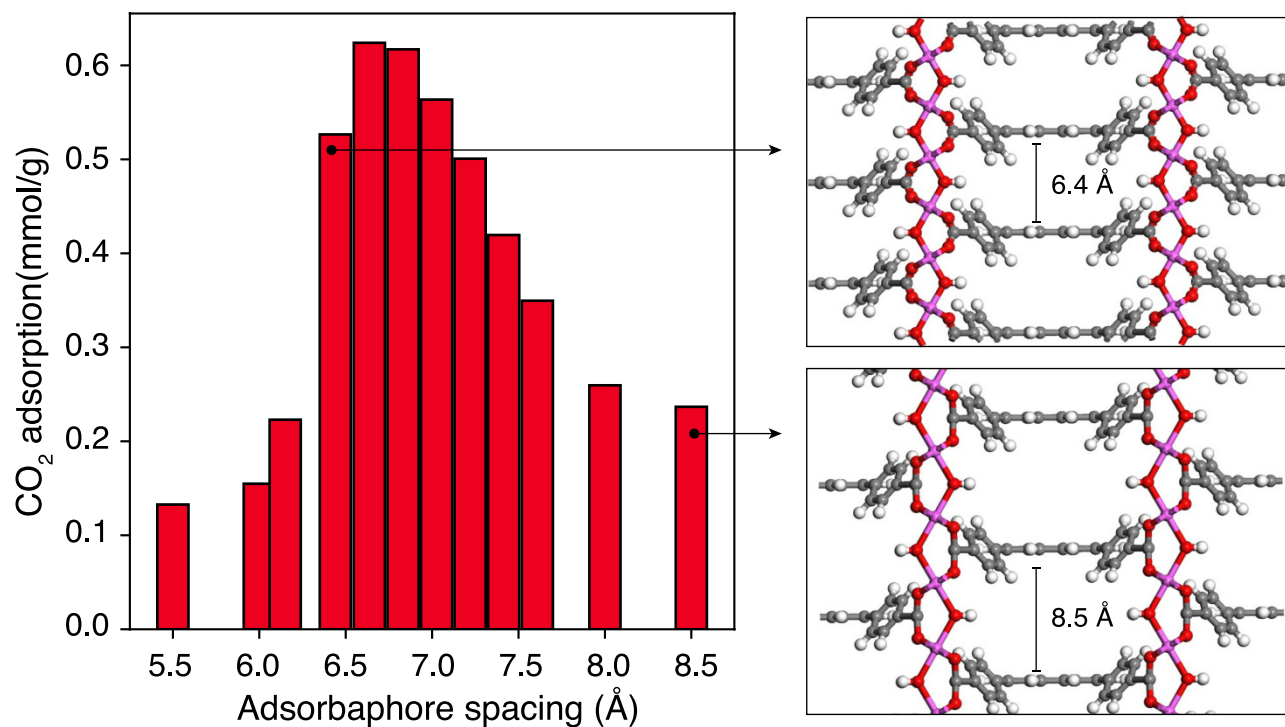
Extended Data Fig. 3 | Examples of adsorbaphores. A selection of the most representative adsorbaphores obtained from visual inspection of the top 50 most frequent adsorbaphores found from the random pairing method described in the Supplementary Information. There are three major trends in the molecular fragments—labelled A1, A2 and A3—which can be observed upon descending each column. The number of chemical features of the fragments increases from left to right across each row. This is accomplished by increasing

the minimum number of common atoms allowed during the substructure search, called the minimum clique threshold (MCT). Pictured in each adsorbaphore is a representative contour map of the energy produced from CO_2 binding with the adsorbaphore atoms from each original CO_2 -binding site. **a–c**, A1, planar aromatic systems, in which CO_2 binds in between the stacked rings; **d–f**, A2, CO_2 binds near the bridging oxygen of a pillared vanadium SBU; **g–i**, A3, CO_2 binds between open-metal Cr SBUs.



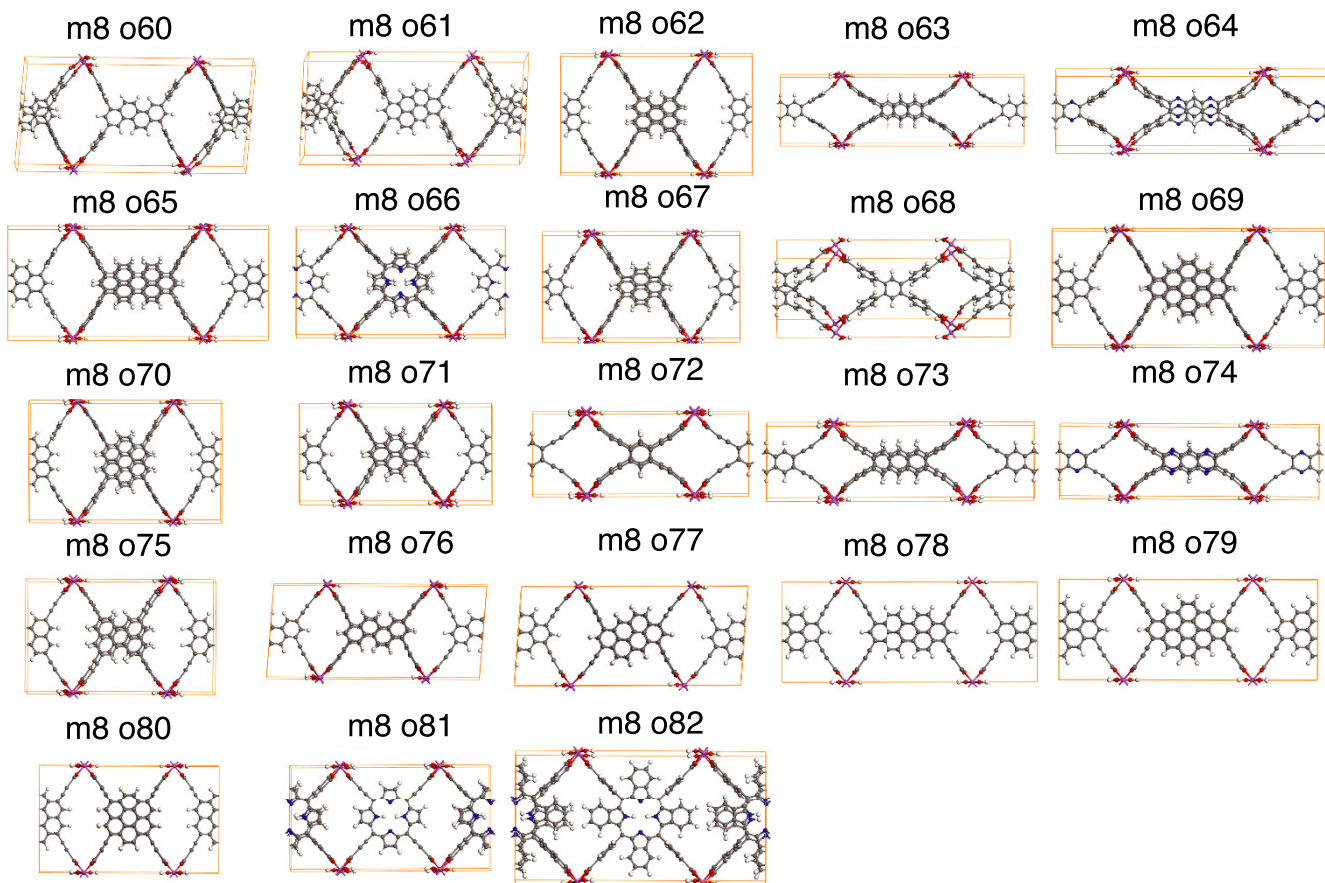
Extended Data Fig. 4 | Frequently found adsorbaphores. a–c, Frequently found adsorbaphores from the maximum clique detection method; A1 (a); A2 (b) and A3 (c). Atom colours: grey, carbon; red, oxygen; green, vanadium; blue,

chromium. d, Representative adsorbaphore A1 found in different hypothetical MOFs from the hypothetical database. The adsorbaphore atoms are highlighted in green.



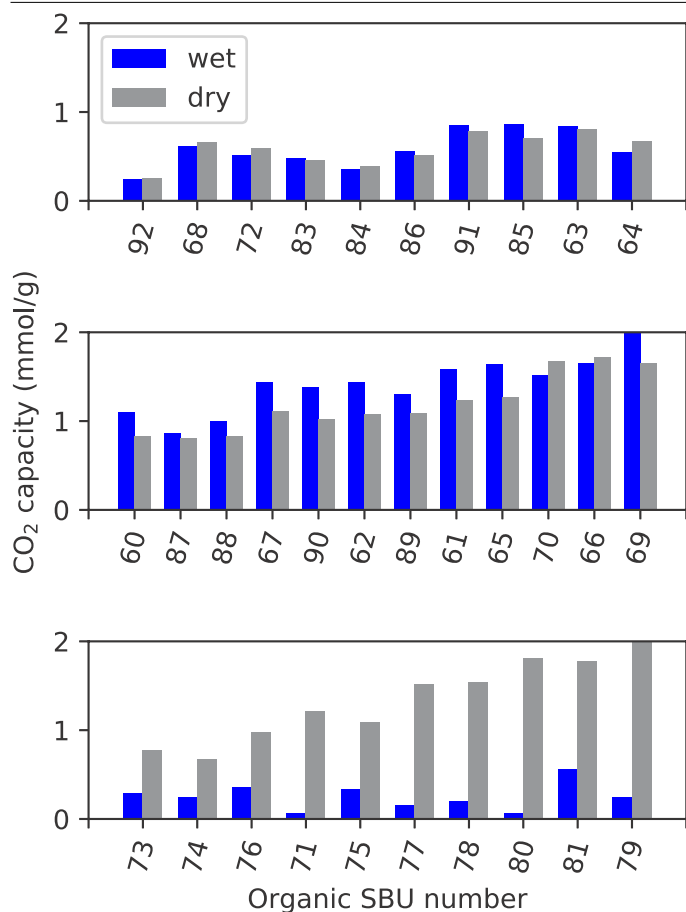
Extended Data Fig. 5 | Effect of the adsorbaphore spacer. Left, plot of the CO₂ adsorption at 0.15 bar and 313 K in hypothetical MOFs with the frz topology, metal node number 8 and organic linker number 67 (m8o67). The interplanar

spacing of the adsorbaphore atoms is adjusted by reassembling the structure with longer or shorter Al–O bonds. Right, representation of the spacing adjustments made to the material.

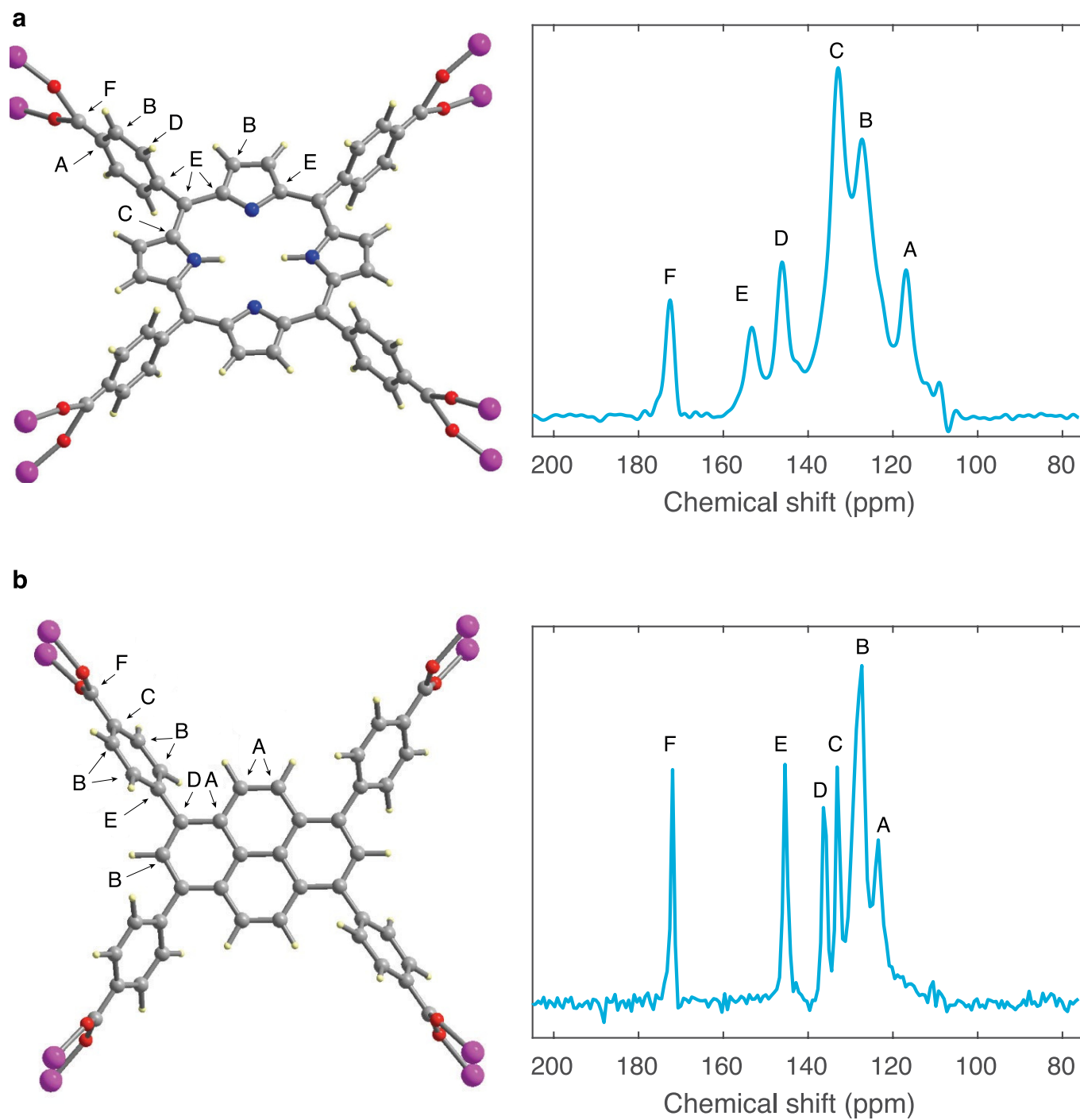


Extended Data Fig. 6 | Hypothetical MOFs built with the frz topology. These structures contain an Al one-dimensional rod (m8) and the organic ligands 60–82 from Extended Data Figs. 1 and 2. The spacing between parallel aromatic cores (seen in the centre of each hypothetical MOF) is around 6.7 Å, defining

the adsorbaphore site in each material. No functional groups were used to decorate these materials. We refer to the synthesized versions of **m8o66** and **m8o67** as **Al-PMOF** ($\text{Al}_2(\text{OH})_2(\text{H}_2\text{TCPP})$) and **Al-PyrMOF** ($\text{Al}_2(\text{OH})_2(\text{TBAPy})$), respectively.



Extended Data Fig. 7 | CO₂ adsorption capacity of a class of frz-based hypothetical MOFs at 0.15 bar and 313K under 'wet' (85% relative humidity) and 'dry' flue gas conditions. Top, middle, hypothetical MOFs in which the organic ligand is connected to the Al ion via 4 benzoate moieties; bottom, hypothetical MOFs in which the organic ligand is connected to the Al metal ion via 4 acetylenic carboxylate moieties. The materials are ranked from lowest adsorbaphore density to highest, and the number on the x axis corresponds to the organic linker number (YY) in m8oYY (see Extended Data Fig. 6).



Extended Data Fig. 8 | NMR spectra. a, b, ^{13}C cross-polarization MAS spectrum of **Al-PMOF (a)** and **Al-PyrMOF (b)** recorded at 9.39 T with sample spinning at 8 kHz; the contact time for the cross-polarization experiment was 2 ms. The

letters labelling the peaks of the spectra correspond to the labels on the carbon atoms in the structures on the left.

Extended Data Table 1 | DFT binding energies of gas particles in the adsorbaphore pocket of each MOF synthesized in this work

MOF	Adsorbaphore Binding Energy (kcal/mol)		
	CO ₂	H ₂ O	N ₂
Al-PMOF	-12.6	-10.4	-7.9
Al-PyrMOF	-10.2	-8.7	-7.5

Extended Data Table 2 | DFT optimized cell parameters of hypothetical MOFs

Metal species	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)
Al(III)	15.84	30.33	6.65
Fe(III)	15.96	30.47	6.83
Ga(III)	15.97	30.56	6.77
In(III)	16.23	30.88	7.27
Sc(III)	16.12	30.70	7.34
Y(III)	16.43	30.72	7.70

The MOFs are built with the frz topology and organic linker number 67 (m8o67) and various trivalent metal species. The *c* axis corresponds to the spacing between aromatic rings in the adsorbaphore.

Global satellite-observed daily vertical migrations of ocean animals

<https://doi.org/10.1038/s41586-019-1796-9>

Received: 27 September 2018

Accepted: 10 October 2019

Published online: 27 November 2019

Michael J. Behrenfeld^{1*}, Peter Gaube², Alice Della Penna^{2,3}, Robert T. O'Malley¹, William J. Burt^{4,5}, Yongxiang Hu⁶, Paula S. Bontempi⁷, Deborah K. Steinberg⁸, Emmanuel S. Boss⁹, David A. Siegel^{10,11}, Chris A. Hostetler⁶, Philippe D. Tortell^{4,12} & Scott C. Doney¹³

Every night across the world's oceans, numerous marine animals arrive at the surface of the ocean to feed on plankton after an upward migration of hundreds of metres. Just before sunrise, this migration is reversed and the animals return to their daytime residence in the dark mesopelagic zone (at a depth of 200–1,000 m). This daily excursion, referred to as diel vertical migration (DVM), is thought of primarily as an adaptation to avoid visual predators in the sunlit surface layer^{1,2} and was first recorded using ship-net hauls nearly 200 years ago³. Nowadays, DVMs are routinely recorded by ship-mounted acoustic systems (for example, acoustic Doppler current profilers). These data show that night-time arrival and departure times are highly conserved across ocean regions⁴ and that daytime descent depths increase with water clarity^{4,5}, indicating that animals have faster swimming speeds in clearer waters⁴. However, after decades of acoustic measurements, vast ocean areas remain unsampled and places for which data are available typically provide information for only a few months, resulting in an incomplete understanding of DVMs. Addressing this issue is important, because DVMs have a crucial role in global ocean biogeochemistry. Night-time feeding at the surface and daytime metabolism of this food at depth provide an efficient pathway for carbon and nutrient export^{6–8}. Here we use observations from a satellite-mounted light-detection-and-ranging (lidar) instrument to describe global distributions of an optical signal from DVM animals that arrive in the surface ocean at night. Our findings reveal that these animals generally constitute a greater fraction of total plankton abundance in the clear subtropical gyres, consistent with the idea that the avoidance of visual predators is an important life strategy in these regions. Total DVM biomass, on the other hand, is higher in more productive regions in which the availability of food is increased. Furthermore, the 10-year satellite record reveals significant temporal trends in DVM biomass and correlated variations in DVM biomass and surface productivity. These results provide a detailed view of DVM activities globally and a path for refining the quantification of their biogeochemical importance.

For decades, airplane-mounted lidar instruments have used the back-scattering of light (bbp) from phytoplankton, zooplankton and small fish to locally map the distribution of these organisms in the water column^{9–11}. In contrast to passive ocean sensors that measure reflected sunlight, lidar uses lasers as a light source and thus has the capacity to measure marine organisms both during the day and at night. When

DVM animals are prominent, their nocturnal invasion of the surface ocean is expected to increase bbp at night compared with what the bbp would be in the absence of these animals. Five ship-based examples¹² of this DVM signature are shown in Fig. 1a, in which the animal signals contribute 7–28% of total night-time bbp. These signals appear as large spikes in this record because DVM animals are 'bright' targets

¹Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA. ²Applied Physics Laboratory, Air-Sea Interaction and Remote Sensing Department, University of Washington, Seattle, WA, USA. ³Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR 6539 CNRS-Ifremer-IRD-UBO-Institut Universitaire Européen de la Mer (IUEM), Plouzané, France. ⁴Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada. ⁵College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, AK, USA. ⁶NASA Langley Research Center, Hampton, VA, USA. ⁷Earth Science Division, Science Mission Directorate, National Aeronautics and Space Administration Headquarters, Washington, DC, USA. ⁸Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, USA. ⁹School of Marine Sciences, University of Maine, Orono, ME, USA. ¹⁰Earth Research Institute, University of California Santa Barbara, Santa Barbara, CA, USA. ¹¹Department of Geography, University of California Santa Barbara, Santa Barbara, CA, USA. ¹²Botany Department, University of British Columbia, Vancouver, British Columbia, Canada. ¹³Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA.

*e-mail: mjb@science.oregonstate.edu

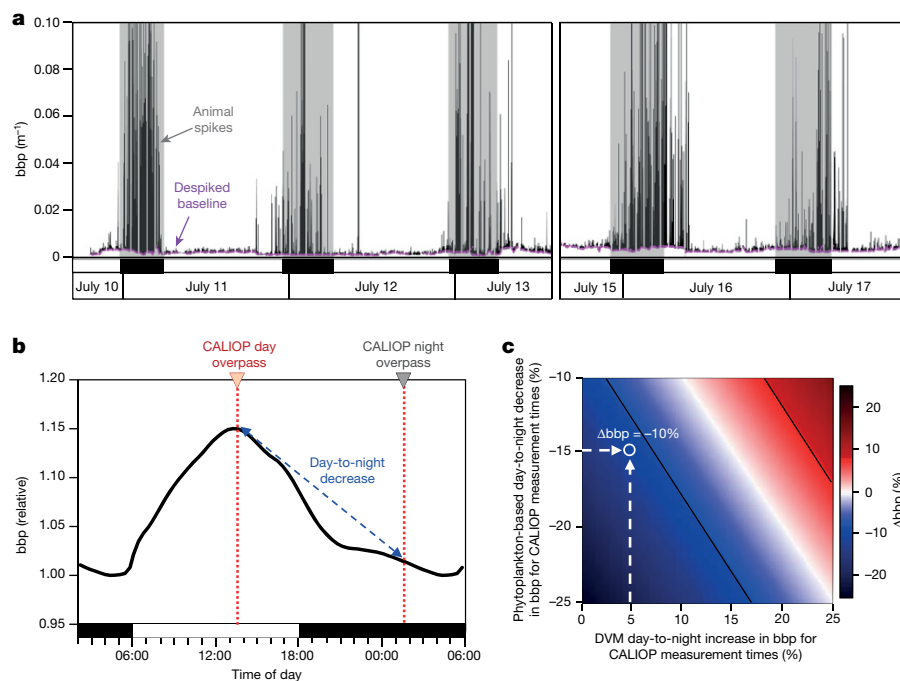


Fig. 1 | Marine animal and phytoplankton influences on day-to-night changes in particulate backscattering coefficients and the biomass-normalized difference ratio. **a**, Ship-based time-series of bbp showing nocturnal spikes associated with DVM animals and the general lack of these spikes during the day¹² (Methods). Black line, complete bbp record. Purple line, despiked baseline bbp record. **b**, Typical diel cycle in bbp of phytoplankton¹⁹ (Methods). Red dotted line, daily equator crossing times of CALIOP. **a, b**, Black and white bars on the x axis indicate night and day, respectively. **c**, Biomass-normalized difference ratios (Δbbp) for ranges in DVM (x axis) and

phytoplankton (y axis) contributions to day-to-night bbp changes. For example, if the phytoplankton diel cycle corresponds (as in **b**) to a 15% day-to-night decrease in bbp at the two CALIOP measurement times (horizontal white dashed arrow) and this decrease is countered by a 5% increase in bbp from DVM animals (vertical white dashed arrow), then Δbbp will have a value of -10% (colour inside labelled white circle). Diagonal black lines correspond to the colour bar range for the CALIOP observations shown in Fig. 2. Note that values on the y axis encompass the range of values expected in the PSO (Extended Data Fig. 5 and Supplementary Discussion).

compared to the phytoplankton and other suspended particles that constitute the much lower bbp baseline¹³ (Fig. 1a); however, animals are also rarer and only occasionally pass through the small sample volume (around 2 ml) of the ship's bbp instrument. By contrast, a lidar with a much larger sampling volume can

effectively capture the signal of all scattering components with every measurement. The satellite cloud-aerosol lidar with orthogonal polarization (CALIOP) sensor has been conducting such measurements for more than a decade^{14,15} (Methods) and its bbp retrievals provide an opportunity to decipher global patterns in vertically migrating animals.

The laser footprint of CALIOP has a diameter of 100 m at the ocean surface and a vertical sampling depth of 22 m in water. What this means is that each CALIOP measurement integrates the bbp signal from a water volume of $1.73 \times 10^5 \text{ m}^3$, which is approximately 5×10^6 greater than the entire water volume measured over a given night in the field data shown in Fig. 1a. The large sampling volume of CALIOP thus ensures that each retrieved bbp value encompasses both the animals and suspended cells and particles in the surface layer. CALIOP is a polar orbiting sensor that conducts daytime and night-time (around 13:40 and 01:40 local time, respectively) near-nadir backscattering measurements along its orbit track at a sampling frequency that is equivalent to every 330 m on the ground. A slightly precessing orbit with a 16-day repeat cycle provides global bbp coverage, but ground tracks oriented in opposite directions on the light and dark sides of Earth mean that day and night bbp samples are rarely spatially coincident within a given 24 h period (Extended Data Fig. 1 and Supplementary Discussion). The global signature of DVM animals was therefore investigated by creating $2^\circ \times 2^\circ$ (latitude by longitude) binned monthly CALIOP daytime (bbp^{day}) and night-time ($\text{bbp}^{\text{night}}$) values of bbp.

In the absence of any DVM, the biomass-normalized bbp difference ratio (equation (1)) is expected to yield a negative value for the measurement times of CALIOP (Fig. 1b):

$$\Delta\text{bbp} = (\text{bbp}^{\text{night}} - \text{bbp}^{\text{day}}) / \text{bbp}^{\text{day}} \quad (1)$$

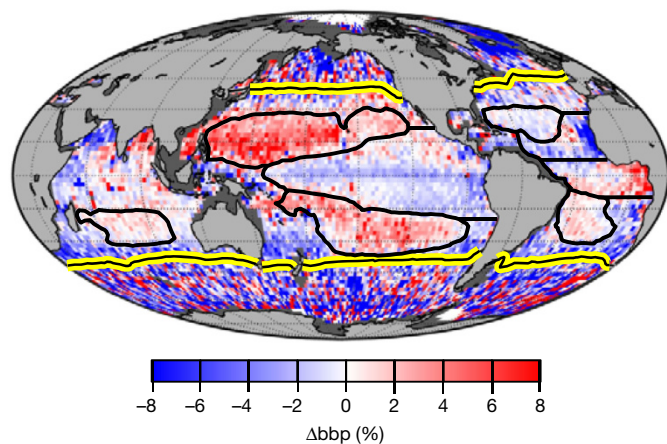


Fig. 2 | Global climatological signal of vertically migrating animals quantified as the normalized difference ratio. Data are $2^\circ \times 2^\circ$ bin means for the 2008–2017 CALIOP record. Yellow-and-black line, contour of annual mean sea surface temperature of 15°C , which effectively separates high-latitude seasonal seas from lower-latitude permanently stratified oceans^{29,30}. Thick black lines separate major regions of the PSO (Fig. 3), in which the 5 subtropical gyres are defined by annual mean surface chlorophyll concentrations of $\leq 0.08 \text{ mg m}^{-3}$. Dark grey, excluded pixels in which the water column depth is less than 1,000 m.

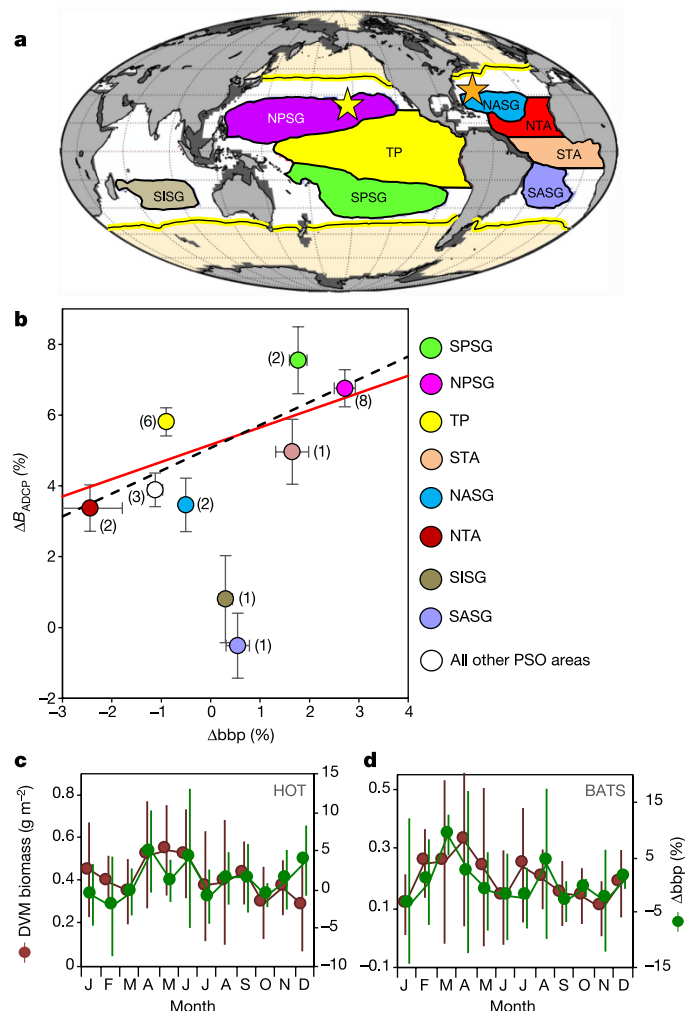


Fig. 3 | Comparison of CALIOP normalized difference ratios and field-based DVM measurements in the PSO. **a**, Colour-coded regions of the PSO from Fig. 2. Yellow star, location of the HOT site. Orange star, location of the BATS site. **b**, CALIOP ($\Delta b b p$) versus field ADCP (ΔB_{ADCp}) normalized difference ratios. Red line, two-sided least-squares linear regression fit to all $5^\circ \times 5^\circ$ bins within our 8 PSO regions (slope = 0.42; F -test, $P = 0.001$, $n = 331$ independent geographical bins) (Extended Data Figs. 7, 8 and Supplementary Discussion). Symbols, regional mean values of $\Delta b b p$ and ΔB_{ADCp} , with s.e.m. shown for each region (SISG, $n = 19$; SASG, $n = 18$; TP, $n = 115$; NASG, $n = 16$; NTA, $n = 23$; NPSG, $n = 59$; STA, $n = 22$; SPSSG, $n = 59$; other, 302). For completeness, the mean value for all PSO bins outside our eight primary regions is indicated by the white symbol. Numbers next to each symbol indicate the median number of days with ADCP data within the 32-year field record for the bins within a given region. The dashed line indicates the two-sided least-squares linear regression fit to regionally averaged data for the $n = 6$ primary regions in which ΔB_{ADCp} is significantly greater than zero (TP, NASG, NTA, NPSG, STA and SPSSG) (slope = 0.64, $r^2 = 0.55$, $P = 0.09$). **c**, Mean monthly field-measured (1994–2005)²¹ migratory zooplankton biomass (g dry weight m^{-2}) for the HOT site (Methods) ($n = 112$ net haul samples) and CALIOP-measured $\Delta b b p$ (2008–2017) ($n = 75$ monthly retrievals) for a $2^\circ \times 2^\circ$ bin centred on the HOT site. Vertical lines indicate 1 s.d. Field and CALIOP data are offset on the x-axis by 6 days to separate s.d. lines. **d**, As in **c**, for field (1994–2017) ($n = 285$ net haul samples) and CALIOP ($n = 87$ monthly retrievals) data for the BATS site.

This is because daytime increases in phytoplankton cell size and organic carbon content and night-time cell division and metabolism^{16–18} create a baseline diel cycle in biomass-normalized bbp with an afternoon maximum and night-time minimum¹⁹ (Fig. 1b and Extended Data Fig. 9). The presence of DVM animals increases bbp^{night} (Fig. 1a)

and therefore causes $\Delta b b p$ to become less negative than the phytoplankton-only signal or even positive if the migrating animals are sufficiently abundant. Thus, phytoplankton and DVM animals together determine $\Delta b b p$ (Fig. 1c), in which the magnitude and sign of $\Delta b b p$ is an index of the DVM signal strength relative to that of the phytoplankton.

The 2008–2017 CALIOP record reveals eight broad regions at tropical and subtropical latitudes (equatorward of the yellow-and-black lines in Fig. 2) where climatological mean $\Delta b b p$ values are spatially coherent. In these regions, plankton populations are relatively stable over time as the water column is permanently stratified within the euphotic zone (referred to hereafter as the permanently stratified ocean (PSO)). These eight broad features (Fig. 2) correspond to the seasonally varying boundaries of the North Pacific (NPSG), South Pacific (SPSSG), North Atlantic (NASG), South Atlantic (SASG) and South Indian (SISG) subtropical gyres, and the higher-nutrient tropical Pacific (TP), north tropical (NTA) and south tropical (STA) Atlantic (Figs. 2, 3a and Extended Data Fig. 2). Within these regions, $\Delta b b p$ values are persistently positive in the NPSG, persistently negative in the TP and NTA, and strongly seasonal in the NASG, STA and all three Pacific areas (Extended Data Fig. 3). In the high-latitude seasonal regions (poleward of the yellow-and-black line in Fig. 2), the CALIOP record reveals large climatological mean values for $\Delta b b p$. However, persistent cloud cover and small-scale spatiotemporal variations in plankton populations in these regions cause within-bin mismatches in bbp^{day} and bbp^{night} data that yield significant bin-to-bin variability in $\Delta b b p$ (Supplementary Discussion). From this point forward, therefore, our analysis primarily focuses on the eight regions of the PSO.

If variability in $\Delta b b p$ is mostly a reflection of the strength of the DVM rather than changes in the phytoplankton cycle, then results shown in Fig. 2 suggest that there are significant regional differences in the relative night-time abundance of these animals. To test this DVM-basis of the regional patterns, we compared CALIOP data to historical field acoustic Doppler current profiler (ADCP) observations of paired day (I_{dB}^{day}) and night (I_{dB}^{night}) acoustic backscatter. Owing to the sparsity of such data, it was necessary to aggregate 32 years (1985–2017) of field measurements into monthly $5^\circ \times 5^\circ$ bins to achieve reasonable global coverage (Methods). We then calculated median values of the normalized difference ratio:

$$\Delta B_{ADCp} = (I_{dB}^{night} - I_{dB}^{day}) / I_{dB}^{day} \quad (2)$$

CALIOP bbp data were then reaggregated into equivalent $5^\circ \times 5^\circ$ bins and regional median values of $\Delta b b p$ were calculated using only those calendar months and bins for which ADCP data were available. Comparison of ΔB_{ADCp} and $\Delta b b p$ for all bins within our eight PSO regions yielded a statistically significant relationship ($P = 0.001$, $n = 331$) (Fig. 3b) with a slope that is very similar to that calculated from regionally averaged values for the six regions in which ADCP data identify a significant DVM signal (Fig. 3b and Supplementary Discussion). In other words, the ADCP and lidar datasets both indicate that DVM animals constitute a greater fraction of night-time plankton communities in the optically clear subtropical gyres that are most advantageous to visual predators. Notably, the regionally averaged data for the SISG and SASG deviate from the other PSO regions (Fig. 3b). The reason for this difference is unknown. These are the only two regions in which ΔB_{ADCp} is near zero, suggesting either that there are few DVM animals or that the influx of night-time DVM animals is compensated by reverse-DVM animals that leave the surface layer²⁰. By contrast, CALIOP data indicate that there is a moderate DVM signal in the SISG and SASG (Fig. 3b). Perhaps the discrepancy in these two regions is simply due to poor ADCP coverage (typically one day–night observation per $5^\circ \times 5^\circ$ bin for the 32-year field record). However, all of the ADCP data in the SASG and all but eight observations in the SISG predate the CALIOP mission, so a temporal change in DVM populations cannot be ruled out.

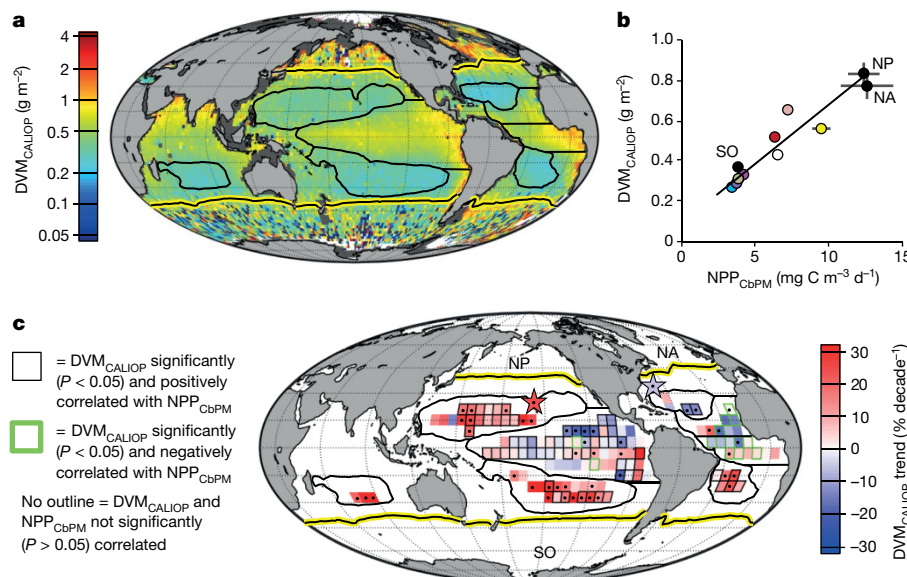


Fig. 4 | CALIOP-based estimates of vertically migrating animal biomass (DVM_{CALIOP}) and temporal changes. a, Annual mean DVM_{CALIOP} (g dry weight m^{-2}). Yellow-and-black and thick black lines as in Fig. 2. **b**, Relationship between annual mean regional DVM_{CALIOP} and mixed-layer net primary production data from the carbon-based production model (NPP_{CbPM} ; $mg\ C\ m^{-3}\ day^{-1}$) (Methods). Solid line, two-sided least-squares linear regression fit ($r^2 = 0.80$; F -test, $P = 0.002$, $n = 9$ geographical regions) of the PSO regions, which are colour-coded according to Fig. 3a. Black symbols, high-latitude North Pacific (NP), North Atlantic (NA) and Southern Ocean (SO) (see c). Horizontal and vertical lines indicate 1 s.d. for annual values between 2008 and 2017 ($n = 111$ months per

geographical region). **c**, Rate of change in DVM_{CALIOP} for $6^\circ \times 6^\circ$ bins for eight PSO regions and at the HOT and BATS sites (coloured stars). Only bins that stay within their respective regional boundaries throughout the year are shown (Extended Data Fig. 2). A black dot in centre of bin indicates that the trend is significant (two-sided least-squares linear regression; F -test, $P < 0.05$, $n = 111$ months per bin). Outlined bins exhibit a significant (two-sided least-squares linear regression; F -test $P < 0.05$, $n = 111$ months per bin) relationship between DVM_{CALIOP} and NPP_{CbPM} , with black and green outlines indicating positive and negative correlations, respectively.

Day and night zooplankton net tows provide quantitative field measurements of DVM biomass, but these labour-intensive measurements are rarely collected at regular intervals over long periods. Two exceptions in the PSO are the sustained records at the Hawaii ocean time-series (HOT) site ($22^\circ 45' N$, $158^\circ W$)²¹ and the Bermuda Atlantic time-series (BATS) site ($31^\circ 40' N$, $64^\circ 10' W$)²² (Fig. 3a). Despite considerable interannual variability, significant ($P < 0.01$) seasonal cycles in DVM biomass (g dry weight m^{-2}) are observed at these locations. For the CALIOP bin centred on the HOT site, the seasonal cycle in Δb_{bp} from January to November is significantly correlated ($r^2 = 0.44$, $P = 0.03$) with field observations (Fig. 3c). Similarly, Δb_{bp} for the CALIOP bin centred on the BATS site exhibits a two-peaked seasonal cycle that correlates well with field observations from August to June ($r^2 = 0.42$, $P = 0.03$) (Fig. 3d). These results are of particular note given the temporal and spatial contrast between datasets (CALIOP data are from $2^\circ \times 2^\circ$ bins collected from 2008 to 2017; zooplankton net data were collected at HOT and BATS from 1994 to 2005²¹ and from 1994 to 2017, respectively). Thus, the correspondence between Δb_{bp} (which includes DVM animals, surface-resident organisms and suspended particles) and animal-specific ADCP²² (Fig. 3b) and net measurements (Fig. 3c, d) gives confidence to the DVM signal detected by CALIOP.

The biomass-normalized Δb_{bp} property reflects the relative strength of the DVM signal, whereas DVM biomass (DVM_{CALIOP}) is more quantitatively related to the simple difference, $b_{bp}^{night} - b_{bp}^{day}$. We estimated DVM_{CALIOP} from this difference as:

$$DVM_{CALIOP} = a(b_{bp}^{night} - c_{bbp}^{day}) \quad (3)$$

in which a is an empirical conversion factor between backscattering (m^{-1}) and biomass ($g\ m^{-2}$) determined from BATS and CALIOP data and c accounts for night–day differences in b_{bp} from growth-rate- and day-length-dependent variations in the diel cycle of phytoplankton (Methods). Application of equation (3) to the CALIOP record yields

a global distribution of DVM_{CALIOP} (Fig. 4a) that differs markedly from Δb_{bp} (Fig. 2). In particular, DVM_{CALIOP} is generally low in the subtropical gyres and high in the nutrient-enriched tropical regions. Therefore, although the relative contribution of DVM animals to night-time plankton communities is higher in clearer waters (Fig. 2), total DVM biomass is higher in more productive regions in which there is a greater availability of food (Fig. 4a). Indeed, the annual mean DVM_{CALIOP} for our PSO regions is highly correlated with passive ocean-colour-based estimates of net primary production ($r^2 = 0.80$, $P = 0.002$) (Fig. 4b, Methods). Results for the high-latitude northern and southern regions are also consistent with this finding (Fig. 4b). Notably, an increase in phytoplankton production yields a less than proportional increase in DVM_{CALIOP} , as might be expected for migrating animals that may be multiple trophic levels removed from the phytoplankton. Consequently, the night–day difference in b_{bp} caused by the diel cycle of phytoplankton (Fig. 1b) generally increases more rapidly than the DVM signal as productivity increases, resulting in negative Δb_{bp} values in productive tropical regions (Fig. 2).

Evidence of long-term changes in zooplankton populations, often linked to climate oscillations, has emerged from field time-series studies^{22–27}. At the HOT site, field-measured DVM zooplankton biomass increased on average by $12.4\ mg\ m^{-2}\ year^{-1}$ (38% per decade) between 1994 and 2005 ($P = 0.04$)²¹. The DVM_{CALIOP} record indicates that this trend continued ($P = 0.05$) at a rate of 23% per decade from 2008 to 2017 (Fig. 4c). At BATS, an overall increasing trend of $7.4\ mg\ m^{-2}\ year^{-1}$ (54% per decade) was reported for field observations from 1994 to 2011 ($P < 0.01$)²². Expanding this dataset to 2017 and re-evaluating the data, we find that DVM biomass increased 63% per decade from 1994 to 2007 ($P = 0.01$) and then decreased 28% per decade from 2008 to 2017 ($P = 0.04$). For this latter period, DVM_{CALIOP} data similarly suggest a decreasing trend (9% per decade; $P > 0.1$) for the $2^\circ \times 2^\circ$ bin that encompasses the BATS site (Fig. 4c). The global coverage provided by CALIOP now allows the evaluation of DVM biomass temporal trends

to be greatly expanded. For example, subdividing our 8 PSO regions (Fig. 3a) into $6^\circ \times 6^\circ$ bins and evaluating only those bins that remain within their respective regional boundaries throughout the year (Extended Data Fig. 2) reveals that decadal trends in DVM_{CALIOF} have coherent geographical patterns (Fig. 4c). Specifically, a predominance of increasing DVM animal biomass is observed in the NPSG, SPSSG, SASG and SISG, whereas decreasing DVM biomass is indicated across much of the tropical regions and the NASG. Moreover, DVM_{CALIOF} is positively correlated with changes in phytoplankton production for most bins, but in the two tropical Atlantic regions the correlations are counter-intuitively inverse (Fig. 4c).

The power of a satellite lidar when studying marine animals lies in its unparalleled annual coverage of the global ocean. By contrast, in this study, it was necessary to compile more than 30 years of ADCP measurements to create a single field-test dataset, and even then spatial and temporal coverage remained poor in many regions (Extended Data Fig. 4). A challenge with satellite lidar data, however, is that the measured night–day bbp differences are not solely owing to DVM animals, so continued work is needed to refine descriptions of the non-DVM contributors (Methods and Supplementary Discussion). Further analyses of field data (for example, ADCP data or continuous plankton recorder survey observations) and modelling²⁸ are needed to fully understand the causative ecological processes that underlie the spatial and temporal DVM patterns that are observed from space, and an advanced satellite lidar with ocean-profiling capabilities¹¹ may contribute new insights into DVM behaviours (Supplementary Discussion). Although there are multiple new avenues to pursue regarding the study of DVM animals, the results presented here provide a step forward in the global exploration of this greatest animal migration on Earth.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1796-9>.

- Lampert, W. The adaptive significance of diel vertical migration of zooplankton. *Funct. Ecol.* **3**, 21–27 (1989).
- Hays, G. C. A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Hydrobiologia* **503**, 163–170 (2003).
- Cuvier, G. *Le Règne Animal distribué d'après son Organisation pour à l'Histoire Naturelle des Animaux et d'Introduction à l'Anatomie Comparée* (Deterville, 1829).
- Bianchi, D. & Mislán, K. A. S. Global patterns of diel vertical migration times and velocities from acoustics data. *Limnol. Oceanogr.* **61**, 353–364 (2016).
- Røstad, A., Kaartvedt, S. & Aksnes, D. L. Light comfort zones of mesopelagic acoustic scattering layers in two contrasting optical environments. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **113**, 1–6 (2016).
- Steinberg, D. K., Goldthwait, S. A. & Hansell, D. A. Zooplankton vertical migration and the active transport of dissolved organic and inorganic nitrogen in the Sargasso Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **49**, 1445–1461 (2002).
- Bianchi, D., Stock, C., Galbraith, E. D. & Sarmiento, J. L. Diel vertical migration: ecological controls and impacts on the biological pump in a one-dimensional ocean model. *Glob. Biogeochem. Cycles* **27**, 478–491 (2013).
- Steinberg, D. K. & Landry, M. R. Zooplankton and the ocean carbon cycle. *Ann. Rev. Mar. Sci.* **9**, 413–444 (2017).
- Churnside, J. H., Wilson, J. J. & Tatarskii, V. V. Lidar profiles of fish schools. *Appl. Opt.* **36**, 6011–6020 (1997).
- Churnside, J. H. & Thorne, R. E. Comparison of airborne lidar measurements with 420 kHz echo-sounder measurements of zooplankton. *Appl. Opt.* **44**, 5504–5511 (2005).
- Hostetler, C. A., Behrenfeld, M. J., Hu, Y., Hair, J. W. & Schullien, J. A. Spaceborne lidar in the study of marine systems. *Ann. Rev. Mar. Sci.* **10**, 121–147 (2018).
- Burt, W. J. & Tortell, P. D. Observations of zooplankton diel vertical migration from high-resolution surface ocean optical measurements. *Geophys. Res. Lett.* **45**, 396–13,404 (2018).
- Briggs, N. T., Slade, W. H., Boss, E. & Perry, M. J. Method for estimating mean particle size from high-frequency fluctuations in beam attenuation or scattering measurements. *Appl. Opt.* **52**, 6710–6725 (2013).
- Behrenfeld, M. J. et al. Space-based lidar measurements of global ocean carbon stocks. *Geophys. Res. Lett.* **40**, 4355–4360 (2013).
- Behrenfeld, M. J. et al. Annual boom–bust cycles of polar phytoplankton biomass revealed by space-based lidar. *Nat. Geosci.* **10**, 118–122 (2017).
- Stramski, D., Shalapyonok, A. & Reynolds, R. A. Optical characterization of the oceanic unicellular cyanobacterium *Synechococcus* grown under a day–night cycle in natural irradiance. *J. Geophys. Res. Oceans* **100**, 13295–13307 (1995).
- DuRand, M. D., Green, R. E., Sosik, H. M. & Olson, R. J. Diel variations in optical properties of *Micromonas pusilla* (Prymniophyceae). *J. Phycol.* **38**, 1132–1142 (2002).
- Dall'Olmo, G. et al. Inferring phytoplankton carbon and eco-physiological rates from diel cycles of spectral particulate beam-attenuation coefficient. *Biogeosciences* **8**, 3423–3439 (2011).
- Kheireddine, M. & Antoine, D. Diel variability of the beam attenuation and backscattering coefficients in the northwestern Mediterranean Sea (BOUSSOLE site). *J. Geophys. Res. Oceans* **119**, 5465–5482 (2014).
- Ohman, M. D., Frost, B. W. & Cohen, E. B. Reverse diel vertical migration: an escape from invertebrate predators. *Science* **220**, 1404–1407 (1983).
- Hannides, C. C. S. et al. Export stoichiometry and migrant-mediated flux of phosphorus in the North Pacific Subtropical Gyre. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **56**, 73–88 (2009).
- Steinberg, D. K., Lomas, M. W. & Cope, J. S. Long-term increase in mesozooplankton biomass in the Sargasso Sea: linkage to climate and implications for food web dynamics and biogeochemical cycling. *Glob. Biogeochem. Cycles* **26**, GB1004 (2012).
- Richardson, A. J. In hot water: zooplankton and climate change. *ICES J. Mar. Sci.* **65**, 279–295 (2008).
- Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A. & Edwards, M. Reorganization of North Atlantic marine copepod biodiversity and climate. *Science* **296**, 1692–1694 (2002).
- Piontkovski, S. A. & Castellani, C. Long-term declining trend of zooplankton biomass in the Tropical Atlantic. *Hydrobiologia* **632**, 365–370 (2009).
- Peterson, W. T. & Schwing, F. B. A new climate regime in northeast Pacific ecosystems. *Geophys. Res. Lett.* **30**, 1896 (2003).
- Chiba, S., Tadokoro, K., Sugisaki, H. & Saino, T. Effects of decadal climate change on zooplankton over the last 50 years in the western subarctic North Pacific. *Glob. Change Biol.* **12**, 907–920 (2006).
- Archibald, K., Siegel, D. A. & Doney, S. C. Modeling the impact of zooplankton diel vertical migration on the carbon export flux of the biological pump. *Glob. Biogeochem. Cycles* **33**, 181–199 (2019).
- Behrenfeld, M. J., Boss, E., Siegel, D. A. & Shea, D. M. Carbon-based ocean productivity and phytoplankton physiology from space. *Glob. Biogeochem. Cycles* **19**, GB1006 (2005).
- Siegel, D. A. et al. Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sens. Environ.* **135**, 77–91 (2013).

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Field data

Data shown in Fig. 1a provide five examples from the subarctic Pacific in which DVM animals significantly increased ship-measured bbp at night. Data are from a previous study¹², which also provides a full description of the study region, measurement system, data processing and uncertainty analyses. The phytoplankton diel cycle in bbp shown in Fig. 1b was calculated from previously published data¹⁹ and is normalized to 1 at the night-time minimum (Supplementary Discussion).

CALIOP data

CALIOP is the primary instrument on the cloud-aerosol lidar and infrared pathfinder satellite observation (CALIPSO) platform³¹, which was launched in 2006. CALIOP measures the total time-dependent return of a pulsed laser output at 1,064 nm and both the co-polarized and cross-polarized return of a 532-nm laser output. The current study is based on night and day differences in the cross-polarized return at 532 nm. The bbp values (available at <http://www.science.oregonstate.edu/ocean.productivity/>) were calculated from this signal following a previously published study¹⁵, except that bbp at 532 nm was not converted to bbp at 440 nm as in the previous study. In the previous study¹⁵, 30° off-nadir CALIOP tilting manoeuvres were conducted every 2 weeks between 2015 and 2016 to collect ocean measurements with minimal surface backscatter from a wide range of ocean environments. Removing molecular backscatter from the CALIOP parallel channel and combining with data from the cross-polarized channel, we then derived particulate backscatter depolarization ratios and compared these data to collocated MODIS diffuse attenuation (K_d) values. This analysis yielded a linear relationship between CALIOP depolarization ratios and the K_d that was used in the previous study¹⁵ to directly retrieve bbp values from CALIOP without reliance on collocated MODIS observations. This same approach was used for the current study. However, we also further evaluated the relationship between particulate depolarization ratios and K_d at 532 nm using an extensive open-ocean airborne lidar dataset compiled from the 2012 Azores campaign¹⁴, the Ship-Aircraft Bio-Optical Research (SABOR) campaign and three campaigns of the North Atlantic Aerosol and Marine Ecosystem Study (NAAMES)³² (no airborne data were collected during the fourth NAAMES campaign as mechanical issues grounded the plane). This airborne-based analysis confirmed the previous CALIOP–MODIS finding of a linear relationship between depolarization ratios and K_d , with a mean depolarization: K_d ratio of 1.76 m and s.d. of 0.19 m. Finally, we calculated bbp using a volume scattering function (VSF) at 180° following a previously published study³³ that was developed using measurements from collocated CALIPSO and MODIS 531-nm data. This VSF is consistent with scattering properties for particles in the phytoplankton size domain and it was applied to both day and night CALIOP data. Organisms larger than phytoplankton, such as DVM animals, will have a VSF with a lower efficiency in the backward direction and enhanced efficiency in the forward direction. This difference is one factor (Supplementary Discussion) that influences the relationship between the CALIOP-retrieved DVM backscatter signal and the biomass of these animals, but it is encompassed in the field-based scaling factor between CALIOP $\text{bbp}^{\text{night}} - \text{bbp}^{\text{day}}$ data and zooplankton dry weight (see below).

CALIOP merges low-gain and high-gain data onboard to reduce data downlink. The onboard gain ratio calculations fail when there are not enough data for both the low- and high-gain channels, which can happen for night-time cross-polarization measurements. At the beginning of the CALIPSO mission, there was an error in the default values of the gain ratios when the onboard calculation fails. This issue rendered the night-time cross-polarization measurements unusable for ocean bbp retrievals until the error was corrected in late 2007, which is why the current study on night–day differences in bbp was limited to the period of 2008–2017. In addition, advanced microwave scanning radiometer

data were used to flag and omit CALIOP retrievals made at wind speeds of $\geq 9 \text{ m s}^{-1}$ to avoid bubble contamination of the bbp values. At wind speeds of $< 9 \text{ m s}^{-1}$, we also applied a depolarization ratio threshold to remove bubble-contaminated data based on statistical analysis from high-wind conditions.

Calculation of $\text{DVM}_{\text{CALIOP}}$

The biomass of DVM animals ($\text{DVM}_{\text{CALIOP}}$) was estimated from CALIOP-measured night–day differences in bbp, an estimate of the day-to-night change in bbp due to the phytoplankton diel cycle alone (Fig. 1b), and a scaling factor between backscattering and zooplankton dry weight.

The value of $\text{bbp}^{\text{night}}$ can be expanded into backscatter by DVM animals (bbp^{DVM}) and backscatter by phytoplankton and other non-migrating particles ($\text{bbp}^{\text{other}}$):

$$\text{bbp}^{\text{night}} = \text{bbp}^{\text{DVM}} + \text{bbp}^{\text{other}} \quad (4)$$

The value of $\text{bbp}^{\text{other}}$ is equal to bbp^{day} corrected for the phytoplankton-based change (c) in bbp from day to night. Rearranging equation (4) and solving for bbp^{DVM} yields:

$$\text{bbp}^{\text{DVM}} = \text{bbp}^{\text{night}} - c\text{bbp}^{\text{day}} \quad (5)$$

A variety of factors may influence the value of c , including the daily division rate of the phytoplankton population, the degree to which this division is synchronized around the day–night cycle, the composition of the phytoplankton community³⁴ and day length. Over much of the PSO, phytoplankton populations are dominated by *Prochlorococcus*, *Synechococcus* and picoeukaryotic species, all of which generally synchronize cell division to the first half of the night^{35–39}. We therefore focused on describing c as a function of both division rate (μ ; divisions per day) and day length.

First, the dependence of c on μ was determined from a previously published laboratory study³⁴ and is described by (Extended Data Fig. 5a and Supplementary Discussion):

$$c_1 = 0.929 - 0.122\mu \quad (6)$$

in which the term -0.122 has units of days per division. Equation (6) was applied to MODIS passive ocean-colour-based estimates of μ calculated for the 2008–2017 period (see below). Second, the dependence of c on day length exists because a change in day length causes the two CALIOP sampling points (determined by its fixed orbit) to line up differently with the phytoplankton diel cycle. For the range of day lengths encountered across the PSO, this day length (dl) dependence is described by (Extended Data Fig. 5b and Supplementary Discussion):

$$c_2 = 0.986 + 0.0012\text{dl} \quad (7)$$

in which dl (h) is known precisely for each CALIOP pixel based on date and latitude and the term 0.0012 has units of h^{-1} .

Monthly DVM zooplankton dry-weight data are available at the BATS site over the 2008–2017 CALIOP period. Comparison of these field data to monthly bbp^{DVM} values calculated from equations (5)–(7) for the $2^\circ \times 2^\circ$ CALIOP bin centred on the BATS site yields a mean scaling factor between bbp^{DVM} and DVM biomass of $1,596 \text{ g m}^{-1}$ (Supplementary Discussion). Applying this conversion factor gives the following expression for $\text{DVM}_{\text{CALIOP}}$:

$$\text{DVM}_{\text{CALIOP}} = 1,596(\text{bbp}^{\text{night}} - c_1 c_2 \text{bbp}^{\text{day}}) \quad (8)$$

In Fig. 3c, d, we compare HOT and BATS DVM zooplankton dry-weight data to the CALIOP-retrieved property, Δbbp , simply to maintain consistency with Fig. 3b (Supplementary Discussion). However, DVM biomass should be quantitatively related to the simple difference,

$\text{bbp}^{\text{night}} - \text{bbp}^{\text{day}}$. The reason that the seasonal cycle in Δbbp corresponds to that of DVM biomass at HOT and BATS is because bbp^{day} varies little over the year at these two sites, making Δbbp highly correlated ($r^2 > 0.99$, $P < 0.001$) with the simple difference, $\text{bbp}^{\text{night}} - \text{bbp}^{\text{day}}$.

Field ADCP data

Global ADCP data were obtained from the Joint Archive for Shipboard ADCP (JASADCP, <http://ilikai.soest.hawaii.edu/sadcp/>). We selected acoustic backscatter data for the upper 20 m of the ocean surface (equivalent to the sampling depth of CALIOP) from ADCPs with frequencies of 150 or 300 kHz and excluded data from frequencies of 38 and 75 kHz. The frequencies of 150 and 300 kHz better isolate scattering from the small (about 0.5–5 mm) animals that are most likely to be detected by CALIOP owing to their much greater abundances. Backscatter amplitude (E) measurements were converted into acoustic intensity (I_{dB}) following a previously published method⁴⁰:

$$I_{\text{dB}} = 10 \log(10^{k_c E/10} - 10^{k_c E_{\text{noise}}/10}) \quad (9)$$

in which k_c is a scaling factor used to convert backscatter amplitude counts to decibels (dB) and E_{noise} is the noise floor of the individual ADCP dataset. Values of k_c are frequency-dependent and were taken from a previous study⁴⁰. The noise floor was defined as the minimum $k_c E$ for each individual data file. A total of 7,622 individual paired diel cycles were extracted from the global database (Extended Data Fig. 4). The geographical distribution of these diel cycle data is largely biased to the regions surrounding the Hawaiian Islands, along the Tropical Ocean atmosphere (TAO/TRITON) mooring array in the equatorial Pacific, the Southern California Bight, between Chile and Antarctica, and off the northeast United States seaboard (Extended Data Fig. 4). Detailed inspection of echograms computed from each individual data file revealed occasional time-synchronicity issues, with the apparent deep DVM occurring more than 2 h before or after local sunrise or sunset. To validate the time stamp of each data file, we compared the timing of the mesopelagic (maximum depth of the data file to 150 m) DVM to the time of local sunrise and sunset. Time-synchronicity issues were detected in less than 10% of the data files. The time stamp of these errant ADCP backscattering observations were corrected to synchronize the deep DVM with sunrise or sunset before computing the day–night differences analysed in the manuscript.

Field time-series zooplankton biomass data

Monthly climatological migratory zooplankton dry-weight data shown in Fig. 3c for HOT were calculated using data from Fig. 1 of a previously published study²¹. Monthly climatological migratory zooplankton dry-weight data shown in Fig. 3d for BATS were calculated from data from 1994 to 2017 provided by D.K.S. The 1994–2011 subset of these data has previously been published²².

Global phytoplankton NPP, biomass, and division rate data

Phytoplankton NPP values used for Fig. 4b, c, phytoplankton division rates (μ) used for calculating c_1 in equation (6) and phytoplankton biomass (C_{phyto}) shown in Extended Data Fig. 6 are from the CbPM⁴¹

using MODIS passive ocean-colour data collected between 2008 and 2017 (data are available at <http://www.science.oregonstate.edu/ocean.productivity/>).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The CALIOP lidar and field ADCP datasets analysed during the current study are available at <http://www.science.oregonstate.edu/ocean.productivity/> and from the Joint Archive for Shipboard ADCP at <http://ilikai.soest.hawaii.edu/sadcp/>. Source Data for Figs. 3b–d, 4b and Extended Data Figs. 3, 6–8 are provided with the paper.

- Winker, D. M. et al. Overview of the CALIPSO mission and CALIOP data processing algorithms. *J. Atmos. Ocean. Technol.* **26**, 2310–2323 (2009).
- Behrenfeld, M. J. et al. The North Atlantic Aerosol and Marine Ecosystem Study (NAAMES): science motive and mission overview. *Front. Mar. Sci.* **6**, 122 (2019).
- Lu, X. et al. Retrieval of ocean subsurface particulate backscattering coefficient from space-borne CALIOP lidar measurements. *Opt. Express* **24**, 29001–29008 (2016).
- DuRand, M. D. & Olson, R. J. Diel patterns in optical properties of the chlorophyte *Nannochloris* sp.: relating individual-cell to bulk measurements. *Limnol. Oceanogr.* **43**, 1107–1118 (1998).
- Vaulot, D., Marie, D., Olson, R. J. & Chisholm, S. W. Growth of *Prochlorococcus*, a photosynthetic prokaryote, in the equatorial Pacific Ocean. *Science* **268**, 1480–1482 (1995).
- André, J. M., Navarette, C., Blanchot, J. & Radenac, M. H. Picophytoplankton dynamics in the equatorial Pacific: growth and grazing rates from cytometric counts. *J. Geophys. Res.* **104**, 3369–3380 (1999).
- Vaulot, D. & Marie, D. Diel variability of photosynthetic picoplankton in the equatorial Pacific. *J. Geophys. Res. Oceans* **104**, 3297–3310 (1999).
- Binder, B. J. & DuRand, M. D. Diel cycles in surface waters of the equatorial Pacific. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **49**, 2601–2617 (2002).
- Jacquet, S., Prieur, L., Avois-Jacquet, C., Lennon, J. F. & Vaulot, D. Short-timescale variability of picophytoplankton abundance and cellular parameters in surface waters of the Alboran Sea (western Mediterranean). *J. Plankton Res.* **24**, 635–651 (2002).
- Gostiaux, L. & van Haren, H. Extracting meaningful information from uncalibrated backscattered echo intensity data. *J. Atmos. Ocean. Technol.* **27**, 943–949 (2010).
- Westberry, T. K., Behrenfeld, M. J., Siegel, D. A. & Boss, E. Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Glob. Biogeochem. Cycles* **22**, GB2024 (2008).

Acknowledgements This work was supported by the National Aeronautics and Space Administration's North Atlantic Aerosol and Marine Ecosystems Study (NAAMES) and EXport Processes in the Ocean from RemoTe Sensing (EXPORTS) study. A.D.P. was supported by the Applied Physics Laboratory Science and Engineering Enrichment Development (SEED) fellowship. This project received funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement number 749591.

Author contributions M.J.B. designed the study. M.J.B., P.G., A.D.P., W.J.B., Y.H. and R.T.O. processed data and analysed results. M.J.B., P.G., A.D.P. and R.T.O. prepared display items. M.J.B. wrote the manuscript with contributions from all authors.

Competing interests The authors declare no competing interests.

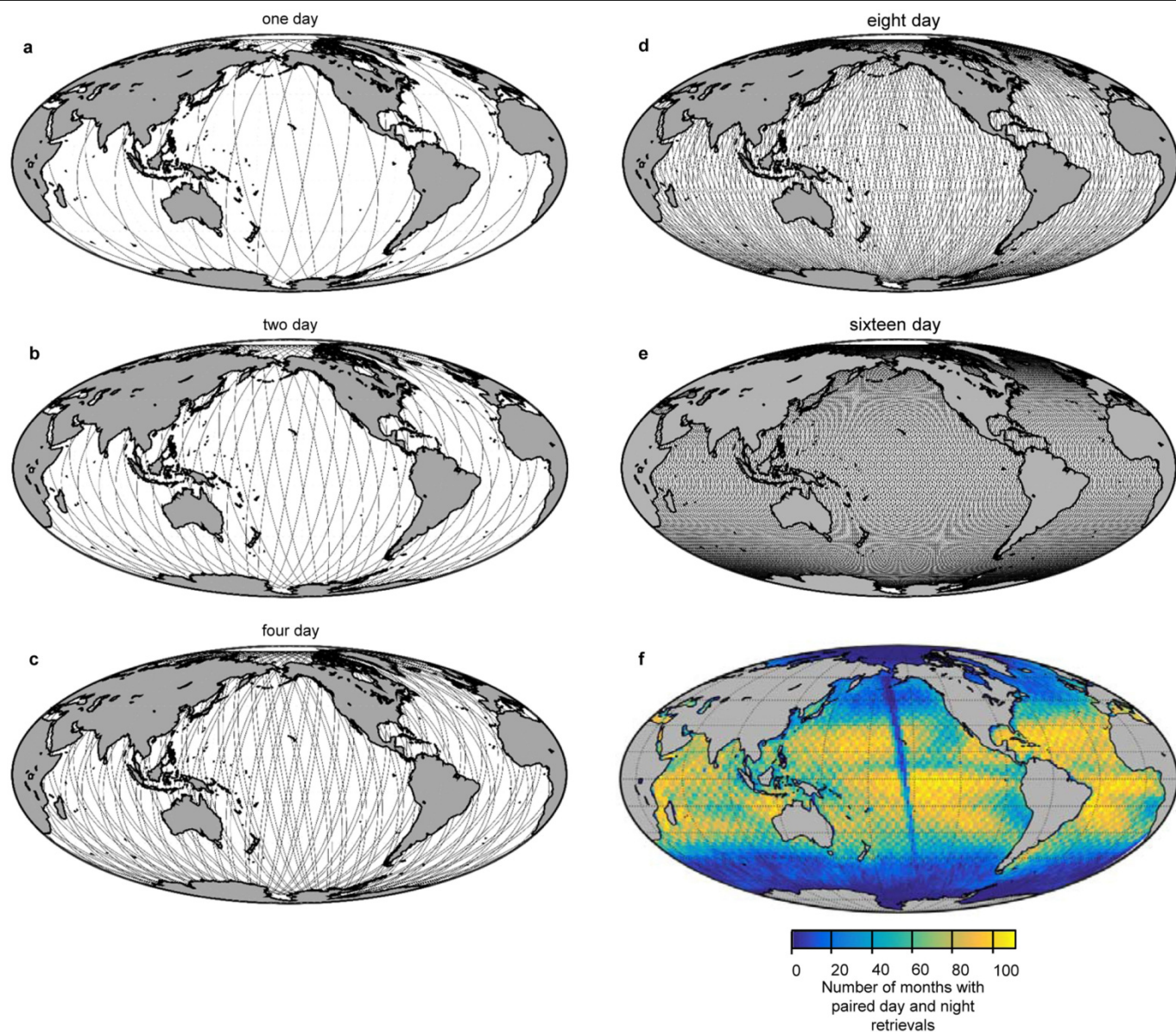
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1796-9>.

Correspondence and requests for materials should be addressed to M.J.B.

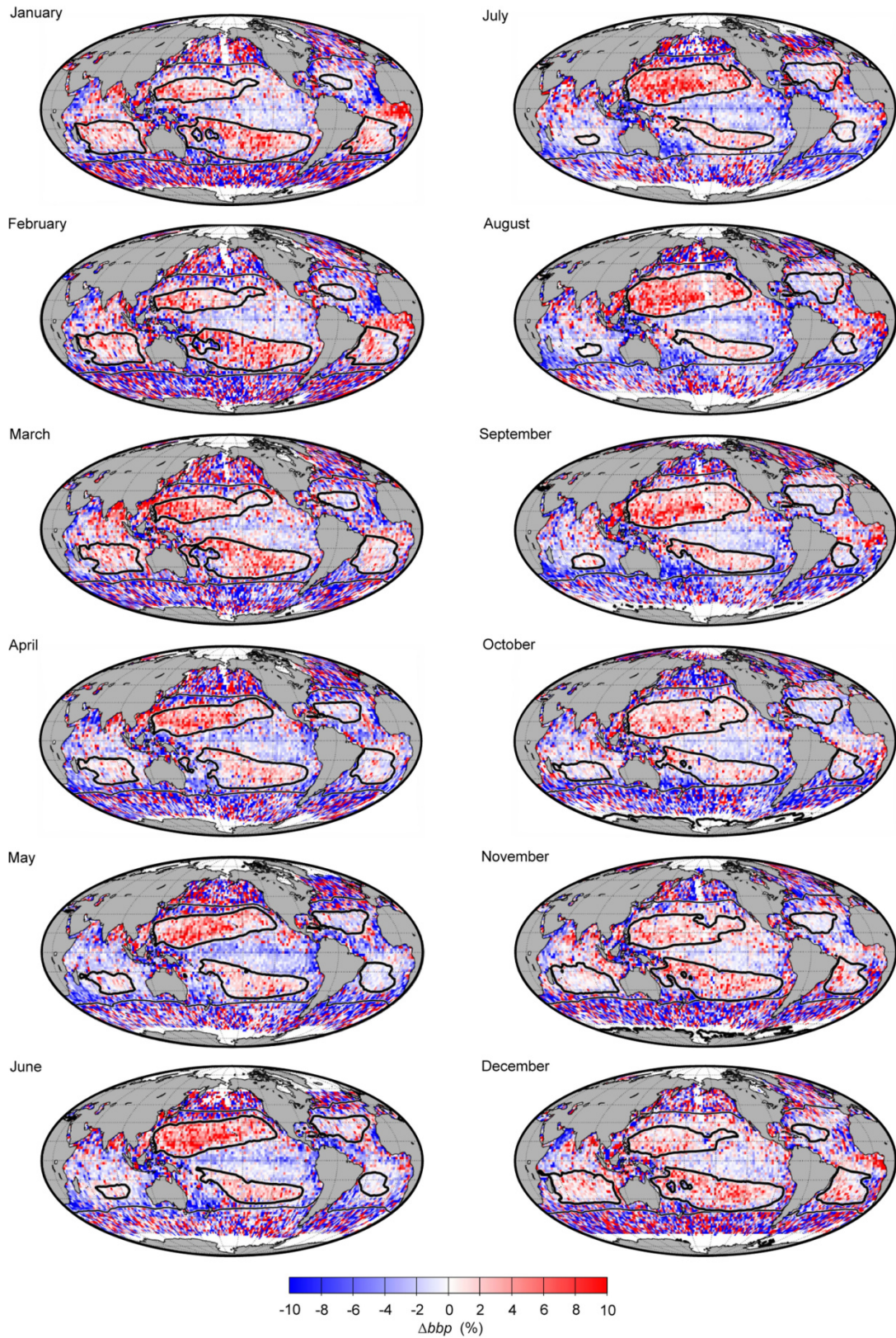
Peer review information Nature thanks Daniele Bianchi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

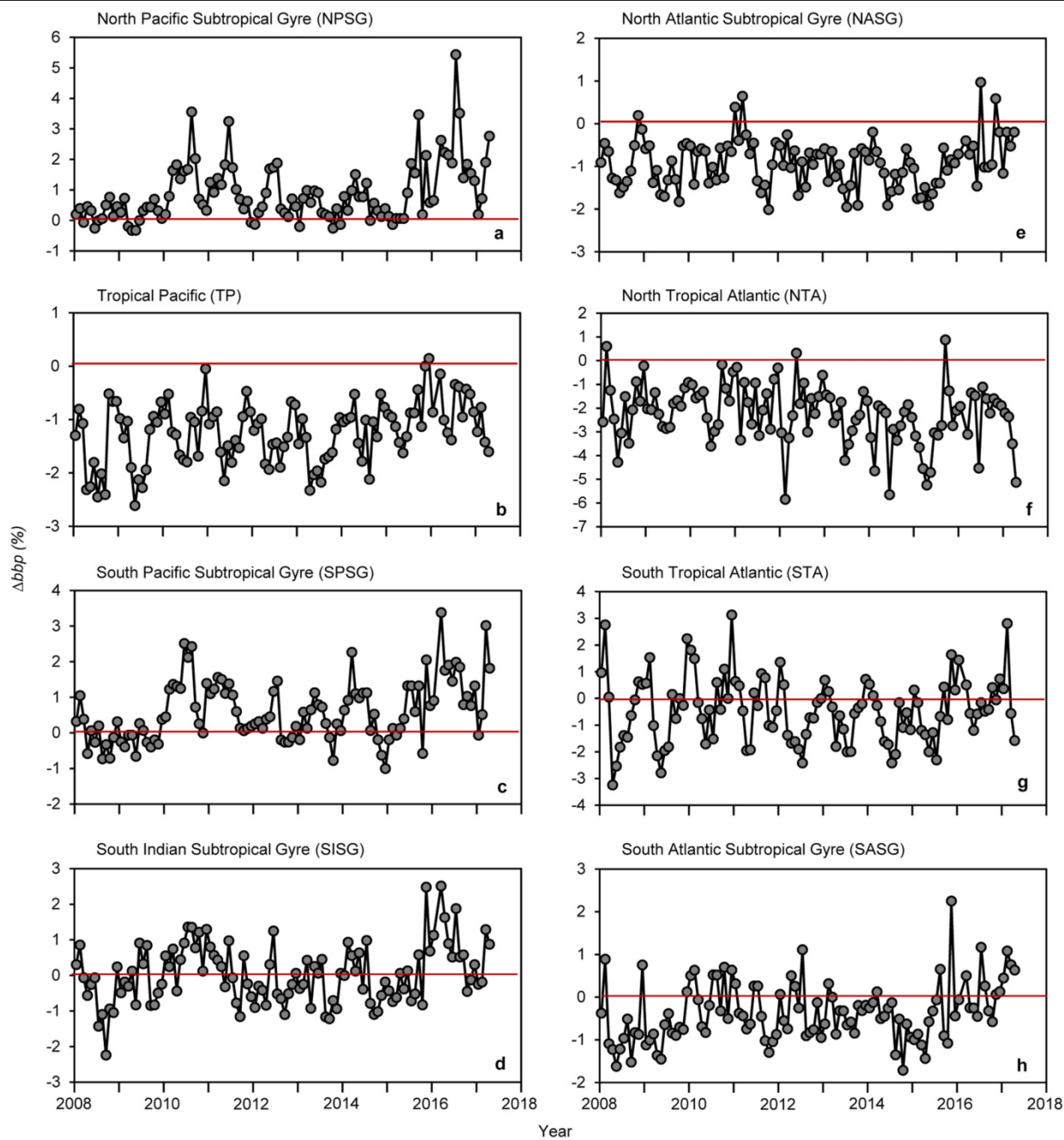


Extended Data Fig. 1 | Global CALIOP observational coverage. a–e, CALIOP ground tracks for 1 (a), 2 (b), 4 (c), 8 (d) and 16 (e) days. f, Number of months for each $2^\circ \times 2^\circ$ bin with day and night retrievals of bbp for the 2008–2017 study period. The total number of months possible is 115. The north–south strip of

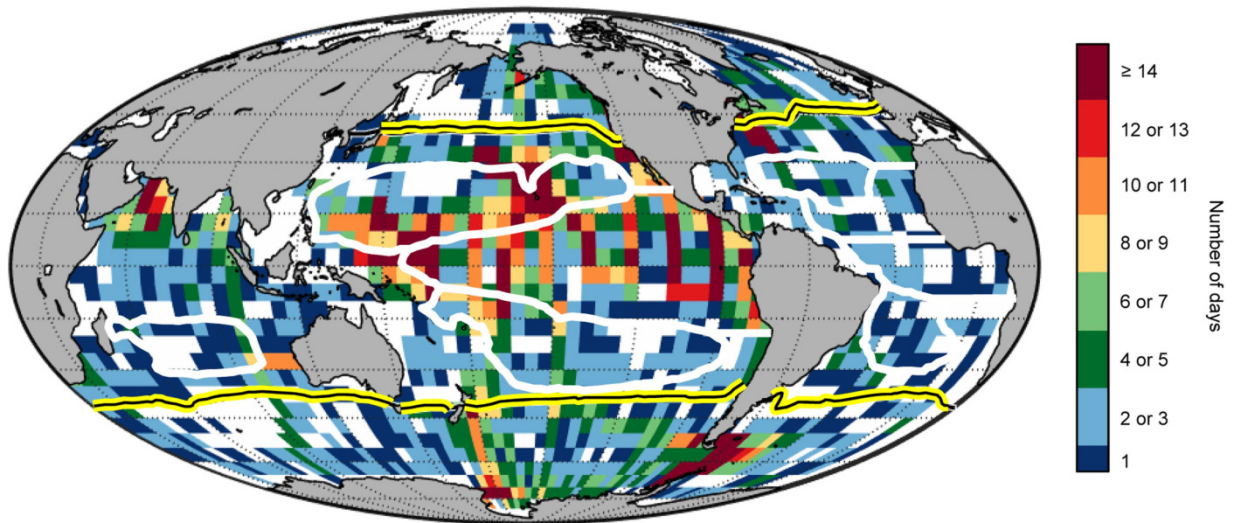
low retrieval success in the middle of the Pacific is caused by a gap in ancillary AMSER surface wind data. AMSER wind data are used for flagging CALIOP data with potential bubble contamination (Methods).



Extended Data Fig. 2 | Global distributions of monthly climatological mean values of Δbbp . Thin black line, contour of monthly mean sea surface temperature of 15°C. Thick black lines, monthly extent of the 5 subtropical gyres in which annual mean surface chlorophyll concentrations are $\leq 0.08 \text{ mg m}^{-3}$.

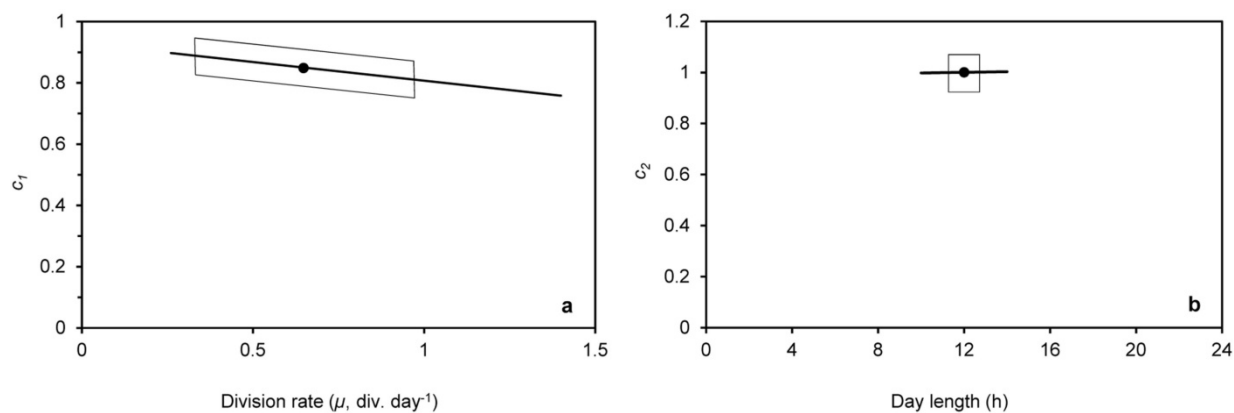


Extended Data Fig. 3 | Time series of $\Delta b b p$ for the PSO. a–h, The 2008–2017 monthly values of $\Delta b b p$ (%) for the eight PSO regions described in Fig. 3a.



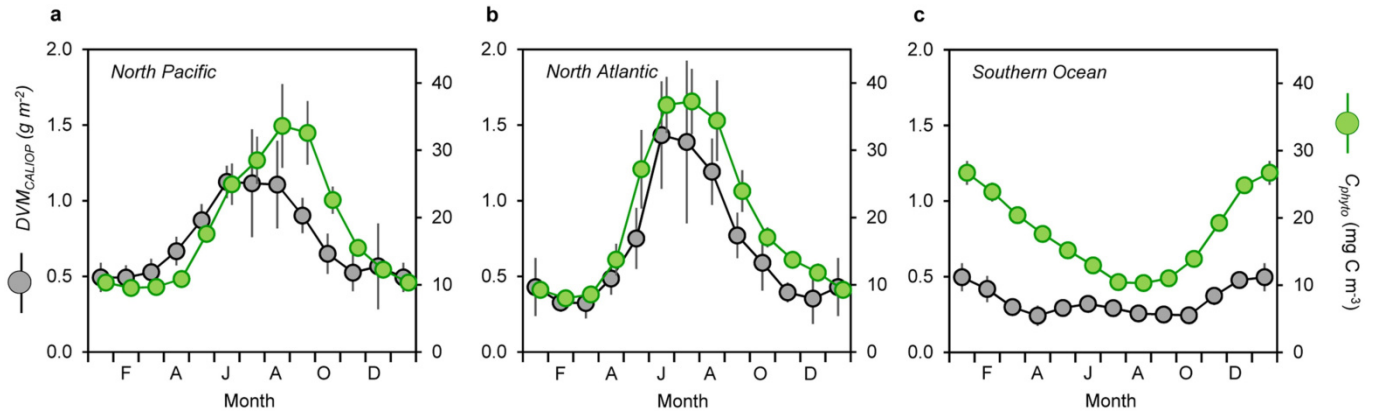
Extended Data Fig. 4 | Global coverage of field ADCP data. Number of days within each $5^\circ \times 5^\circ$ bin for which paired day–night ADCP data are available from the 1985–2017 JASADCP-based field archive (<http://ilikai.soest.hawaii.edu/>

sadcp/). The total number of days possible is 11,680. White bins, no data. Yellow/black line, contour of annual mean sea surface temperature of 15°C . Thick white lines, boundaries of the eight PSO regions described in Fig. 3a.



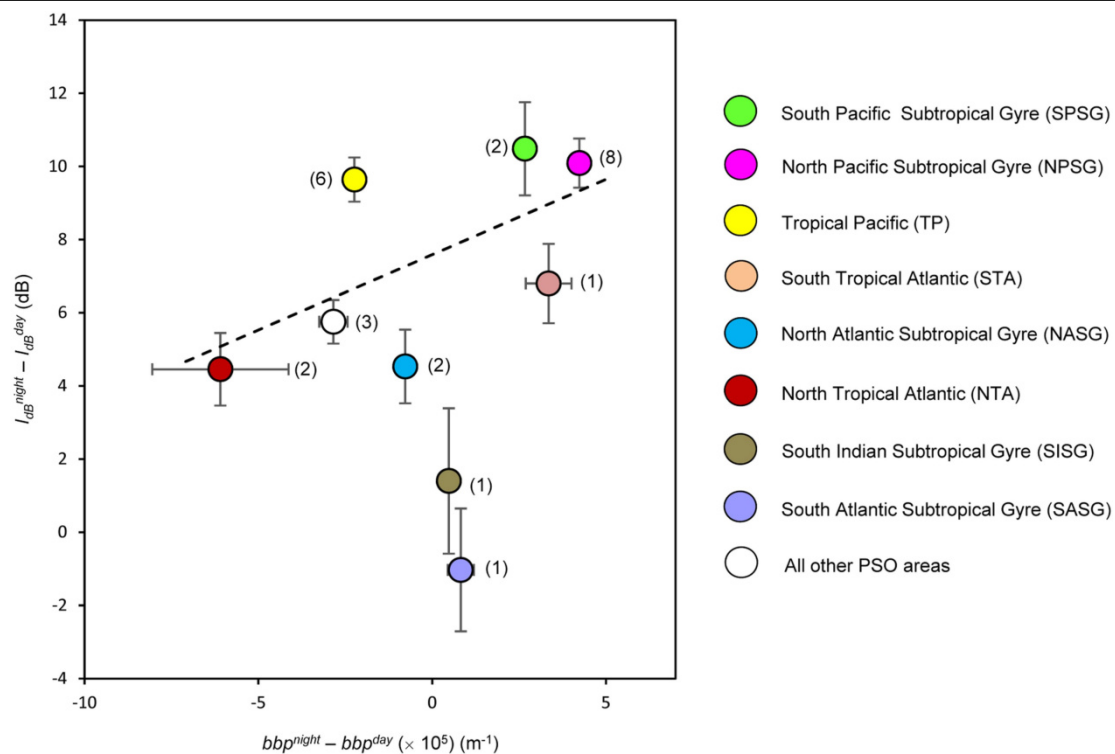
Extended Data Fig. 5 | Influence of phytoplankton division rate and of day length on calculated DVM backscatter for the PSO. a. Values for c_1 (equation (6)) over the range of phytoplankton division rates (μ) in the PSO ($n = 999$ monthly μ values for all PSO regions). Solid circle, mean value of μ and c_1 for the PSO. The box shows ± 1 s.d. of the mean of μ and the solid line shows

values of c_1 over the full range in μ for the PSO. **b.** Values for c_2 (equation (7)) over the range of day lengths in the PSO ($n = 999$ monthly day length values for all PSO regions). Solid circle, mean day length and c_2 value for the PSO. The box shows ± 1 s.d. of the mean day length and the solid line shows values of c_2 over the full range in day length for the PSO.



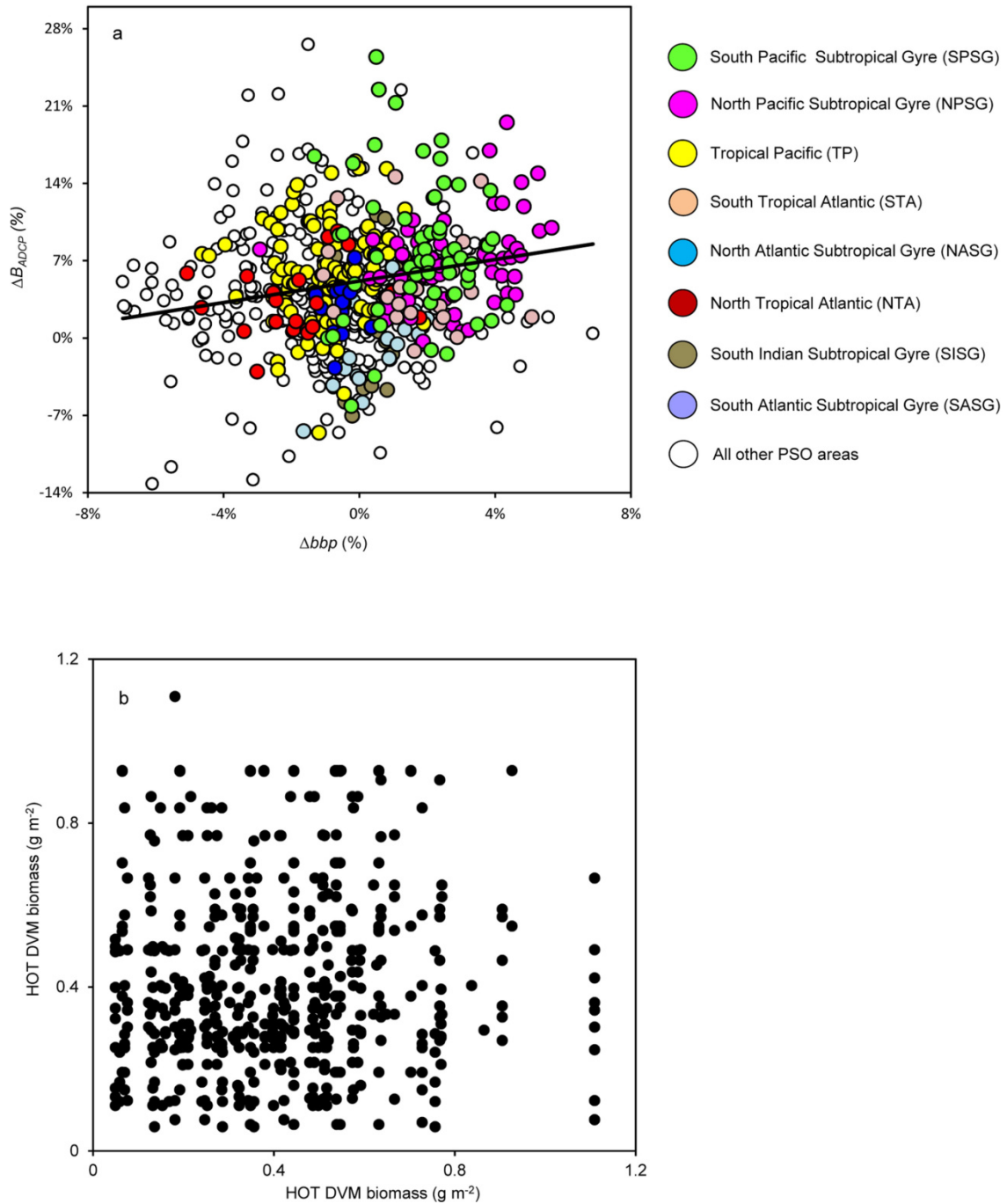
Extended Data Fig. 6 | Seasonal cycles in monthly mean regionally integrated values of DVM_{CALIOP} and phytoplankton biomass for high-latitude regions. a, North Pacific. b, North Atlantic. c, Southern Ocean. These three regions are described in Fig. 4c. Vertical lines show ± 1 s.d. ($n = 111$ monthly

DVM_{CALIOP} (g m⁻²) and C_{phyto} (mg C m⁻³) values for each region). C_{phyto} data are from the carbon-based production model (CbPM) and MODIS passive ocean-colour data (Methods).



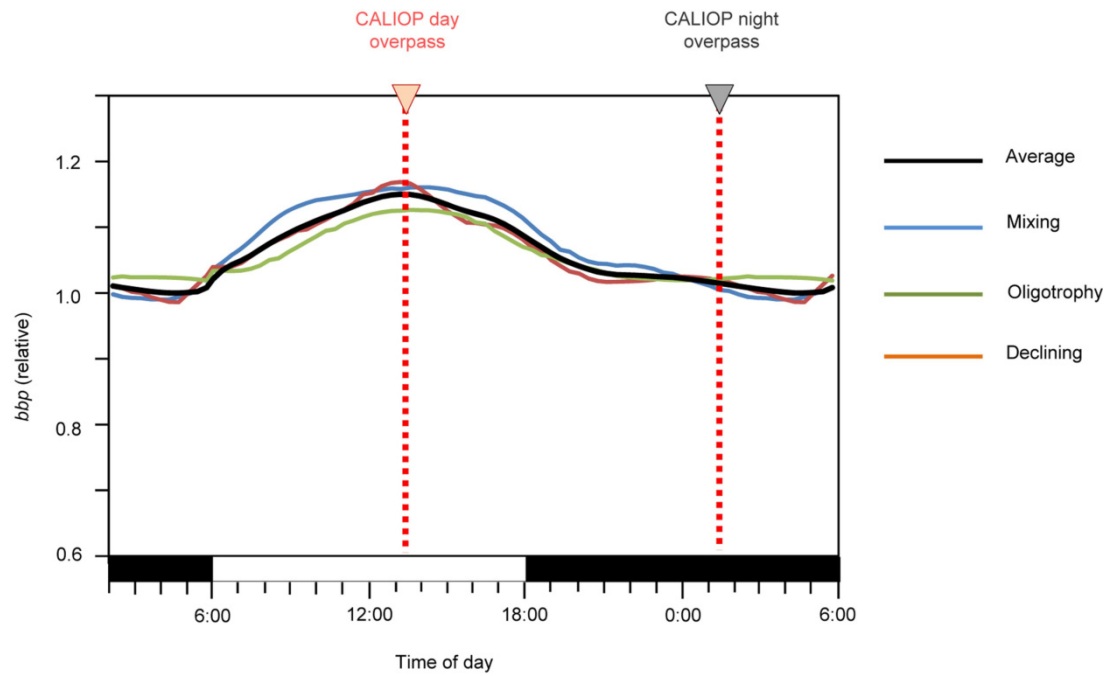
Extended Data Fig. 7 | Comparison of CALIOP night-day bbp differences and field ADCP night-day differences in acoustic backscatter. Dashed line, two-sided least-squares linear regression fit to data for the SPSG, NPSG, TP, STA, NASG and NTA ($n=6$). For completeness, the mean value for PSO bins outside our eight primary regions is indicated by the white symbol. Symbols, regional

mean \pm s.e.m. (SISG, $n=19$; SASG, $n=18$; TP, $n=115$; NASG, $n=16$; NTA, $n=23$; NPSG, $n=59$; STA, $n=22$; SPSG, $n=59$; other, $n=302$). Symbol colours identify region (labelled on the right) and correspond to the colours shown in Fig. 3a. Numbers next to each symbol indicate the median number of days with ADCP data within the 32-year field record for the $5^\circ \times 5^\circ$ bins.



Extended Data Fig. 8 | Bin-to-bin comparison of CALIOP Δbbp and field-based DVM measurements in the PSO. **a**, CALIOP normalized difference ratios (Δbbp) versus field ADCP (ΔB_{ADCP}) normalized difference ratios for $5^\circ \times 5^\circ$ bins within the PSO. Black line, two-sided least-squares linear regression fit (F -test P value for slope; $P < 0.001$; $n = 331$ independent geographical bins) for all data from our eight primary PSO regions (coloured symbols are labelled on the

right). White symbols, PSO values for $5^\circ \times 5^\circ$ bins outside of the 8 primary regions. Inclusion of these data in the linear regression analysis increases the F -test value to $P = 0.005$ ($n = 633$ independent geographical bins). **b**, Relationship between field DVM biomass at the HOT site measured for a given calendar month and year (x axis) versus DVM biomass measured during all other years for the same calendar month (y axis).



Extended Data Fig. 9 | Field-based diel cycles in bbp. Mean diel cycles in bbp from a previous study¹⁹ for mixing (blue line; $n = 69$ days of measurements), oligotrophy (green line; $n = 322$ days of measurements) and declining (red line;

$n = 32$ days of measurements) conditions and the mean of these three cycles (black line), which corresponds to the diel cycle in Fig. 1b.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

CALIO data were collected using custom software developed as part of the CALYPSO satellite mission

Data analysis

Data were analyzed using MatLab Version R2014b and Microsoft Excel 2016 MSO (16.0.4849.1000) 32-bit

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The CALIO lidar and field ADCP datasets analyzed during the current study are available at <http://www.science.oregonstate.edu/ocean.productivity/> and from the Joint Archive for Shipboard ADCP at <http://ilikai.soest.hawaii.edu/sadcp/>. Data presented in figures 3b-d and 4b and Extended Data figures 3, 6, 7, and 8 were provided as spate tabs in an excel file as part of this manuscript submission.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences

☐ Behavioural & social sciences

☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Global CALIOP satellite lidar data were collected continuously over a 10 year period at 330 m sampling distance between retrievals, which are only made along the satellite orbit track (a detailed description of CALIOP data coverage is provided in the Supplementary Discussion for the manuscript). These data were compared to historical ADCP field measurements of acoustic backscatter, with a total of 7,622 day-night ADCP measurement pairs. CALIOP data were also compared to field zooplankton measurements at the long-term HOT and BATS field sampling sights, with a total of 397 zooplankton collections for the two sites.

Research sample

Global CALIOP satellite data shown in figure 2 of the main manuscript were binned into 2 degree latitude by 2 degree longitude bins to provide high signal-to-noise retrievals at high spatial resolution. For the CALIOP-ADCP comparison, CALIOP data were binned into 5 degree latitude by 5 degree longitude bins to be comparable with the much coarser resolution of field ADCP data. For figure 4c of the manuscript, CALIOP data were binned into 6 degree latitude by 6 degree longitude bins to increase signal-to-noise for detecting temporal trends in the CALIOP record. CALIOP data analyzed during the current study are available at <http://www.science.oregonstate.edu/ocean.productivity/> and are reflective of optical properties in the upper 22 m of the ocean surface. For ADCP data, all available field measurements from the Joint Archive for Shipboard ADCP (JASADCP, <http://ilikai.soest.hawaii.edu/sadcp/>) were obtained and then only those measurements conducted at frequencies of 150 or 300 kHz were used for comparison with CALIOP retrievals because these frequencies better isolate scattering from the small (~0.5 – 5 mm) animals that are most likely detected by CALIOP. ADCP measurements provide depth-resolved information on acoustic backscatter. HOT and BATS field data were taken from the publications from Hannides, C.C.S, et al. (2009) Deep-Sea Res. I. 56, 73-88 and Steinberg, D.K., et al. (2012) Global Biogeochem. Cycl. 26, GB1004, doi:10.1029/2010GB004026, respectively. Sampling frequency was dictated by the protocols of the respective programs and DVM biomass is determined by the difference in zooplankton abundance between day and night net collections.

Sampling strategy

No statistical method was used to determine CALIOP sampling, as the frequency of sampling is determined by the hardware of the lidar instrument. For the current study, the native resolution of CALIOP data is far greater than the analysis bins used for assessing DVM patterns and this high sampling resolution improves the signal-to-noise of the retrieved properties. For ADCP data, no statistical method was used to determine sampling frequency as the native resolution of these data is far greater than the single day-night differences used to assess the DVM signal. HOT and BATS data were taken from the published literature and the multi-year monthly sampling frequency of these data (dictated by predefined program protocols) typically provide approximately 10 or more measurements of DVM biomass for each month from which the mean annual cycle was calculated for the current analysis.

Data collection

Optical signals were obtained using the CALIOP satellite lidar. ADCP data were collected by a diversity of researchers on historical ship transects. HOT and BATS data are previously published, as noted above, and are products of the long-term HOT and BATS programs

Timing and spatial scale

CALIOP data were collected between 2008 and 2017 at a native spatial resolution of 330 m sampling distance between retrievals along the satellite orbit track and then combined into various sized bins as described above. ADCP data were collected between 1985 and 2017 and then all available data were globally combined into 5 degree latitude by 5 degree longitude geographic bins. HOT DVM samples were collected between 1994 and 2005 using a plankton net with a 1 square meter opening at the location indicated in figure 3a of the manuscript. BATS data were collected between 1994 and 2017 using a plankton net with a 0.8 m x 1.2 meter opening at the location indicated in figure 3a of the manuscript.

Data exclusions

Global CALIOP data were initially analyzed and then subsampled for data within the PSO. Higher latitude data were excluded for the CALIOP-ADCP comparison in figure 3 because poor sampling coverage and high spatial variability in plankton properties at these latitudes resulted in noisy day-night comparison results (see detailed discussion in Supplementary Discussion). CALIOP exclusion criteria were not pre-established. ADCP data were subsampled to only include measurement frequencies of 150 or 300 kHz because these frequencies better isolate scattering from the small (~0.5 – 5 mm) animals that are most likely detected by CALIOP. Thus, exclusion criteria were pre-established. These ADCP data were then quality controlled as described in the Methods section. All HOT and BATS data reported in the Hannides, C.C.S, et al. and Steinberg, D.K., et al. manuscripts cited above were included.

Reproducibility

Validation of the CALIOP results was achieved through comparison with historical global ADCP data and zooplankton net data from BATS and HOT. Successful replication was achieved for all regions and months by combining data from multiple years of CALIOP, ADCP, and zooplankton net measurements

Randomization

The 8 PSO regions evaluated were defined first for the 5 gyres using the stated chlorophyll criterion in the manuscript and then the 3 remaining equatorial regions based as the region spanning between the north and south gyres. In the Pacific a single equatorial region was defined and in the Atlantic the equatorial region was split into northern and southern sections

Blinding

Our study did not require blinding because it did not involve experiments

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Enamel proteome shows that *Gigantopithecus* was an early diverging pongine

<https://doi.org/10.1038/s41586-019-1728-8>

Received: 21 June 2019

Accepted: 3 October 2019

Published online: 13 November 2019

Frido Welker^{1*}, Jazmín Ramos-Madrigal¹, Martin Kuhlwiilm², Wei Liao^{3,4}, Petra Gutenbrunner⁵, Marc de Manuel², Diana Samodova⁶, Meaghan Mackie^{1,6}, Morten E. Allentoft⁷, Anne-Marie Bacon⁸, Matthew J. Collins^{1,9}, Jürgen Cox⁵, Carles Lalueza-Fox², Jesper V. Olsen⁶, Fabrice Demeter^{7,10}, Wei Wang^{11*}, Tomas Marques-Bonet^{2,12,13,14*} & Enrico Cappellini^{1*}

Gigantopithecus blacki was a giant hominid that inhabited densely forested environments of Southeast Asia during the Pleistocene epoch¹. Its evolutionary relationships to other great ape species, and the divergence of these species during the Middle and Late Miocene epoch (16–5.3 million years ago), remain unclear^{2,3}. Hypotheses regarding the relationships between *Gigantopithecus* and extinct and extant hominids are wide ranging but difficult to substantiate because of its highly derived dentognathic morphology, the absence of cranial and post-cranial remains^{1,3–6}, and the lack of independent molecular validation. We retrieved dental enamel proteome sequences from a 1.9-million-year-old *G. blacki* molar found in Chuifeng Cave, China^{7,8}. The thermal age of these protein sequences is approximately five times greater than that of any previously published mammalian proteome or genome. We demonstrate that *Gigantopithecus* is a sister clade to orangutans (genus *Pongo*) with a common ancestor about 12–10 million years ago, implying that the divergence of *Gigantopithecus* from *Pongo* forms part of the Miocene radiation of great apes. In addition, we hypothesize that the expression of alpha-2-HS-glycoprotein, which has not been previously observed in enamel proteomes, had a role in the biomineralization of the thick enamel crowns that characterize the large molars in *Gigantopithecus*^{9,10}. The survival of an Early Pleistocene dental enamel proteome in the subtropics further expands the scope of palaeoproteomic analysis into geographical areas and time periods previously considered incompatible with the preservation of substantial amounts of genetic information.

G. blacki is an extinct, potentially giant hominid species that once inhabited Asia. It was first discovered and identified by von Koenigswald in 1935, when he described an isolated tooth that he found in a Hong Kong drug-store¹¹. The entire *G. blacki* fossil record, dated between the Early Pleistocene (about 2 million years ago (Ma)) and the late Middle Pleistocene (about 0.3 Ma)¹², includes thousands of teeth and four partial mandibles from subtropical Southeast Asia^{1,13,14}. All the locations at which *G. blacki* remains have been found are in or near southern China, stretching from Longgupo Cave, just south of the Yangtze River, to the Xinchong Cave on Hainan Island and, possibly, into northern Vietnam and Thailand^{15,16}.

To address the evolutionary relationships between *Gigantopithecus* and extant hominoids, we performed protein extractions on dentine and enamel samples of a single molar (CF-B-16) found in Chuifeng Cave, China (Extended Data Figs. 1, 2), that has been morphologically assigned to *G. blacki*^{7,8}. The site has been dated, using multiple approaches, to 1.9 ± 0.2 Ma. We processed enamel and dentine samples using recently established digestion-free protocols that were optimized for extremely degraded ancient proteomes¹⁷ (see Methods). Enamel demineralization was replicated using two acids, trifluoroacetic acid (TFA) and hydrochloric acid (HCl).

¹Evolutionary Genomics Section, Globe Institute, University of Copenhagen, Copenhagen, Denmark. ²Institute of Evolutionary Biology (UPF-CSIC), University Pompeu Fabra, Barcelona, Spain. ³School of Earth Sciences, China University of Geosciences, Wuhan, China. ⁴Anthropology Museum of Guangxi, Nanning, China. ⁵Computational Systems Biochemistry, Max Planck Institute of Biochemistry, Martinsried, Germany. ⁶Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark. ⁷Lundbeck Foundation GeoGenetics Centre, Globe Institute, University of Copenhagen, Copenhagen, Denmark. ⁸CNRS FRE 2029 BABEL, Université Paris Descartes, Faculté de Chirurgie Dentaire, Paris, France. ⁹Department of Archaeology, University of Cambridge, Cambridge, UK. ¹⁰UMR7206 Eco-anthropologie, Muséum national d'Histoire naturelle, Musée de l'Homme, Paris, France. ¹¹Institute of Cultural Heritage, Shandong University, Qingdao, China. ¹²Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain. ¹³Centre for Genomic Regulation (CNAG-CRG), Barcelona Institute of Science and Technology, Barcelona, Spain. ¹⁴Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. ¹⁵e-mail: frido.welker@bio.ku.dk; wangw@sdu.edu.cn; tomas.marques@upf.edu; ecappellini@bio.ku.dk

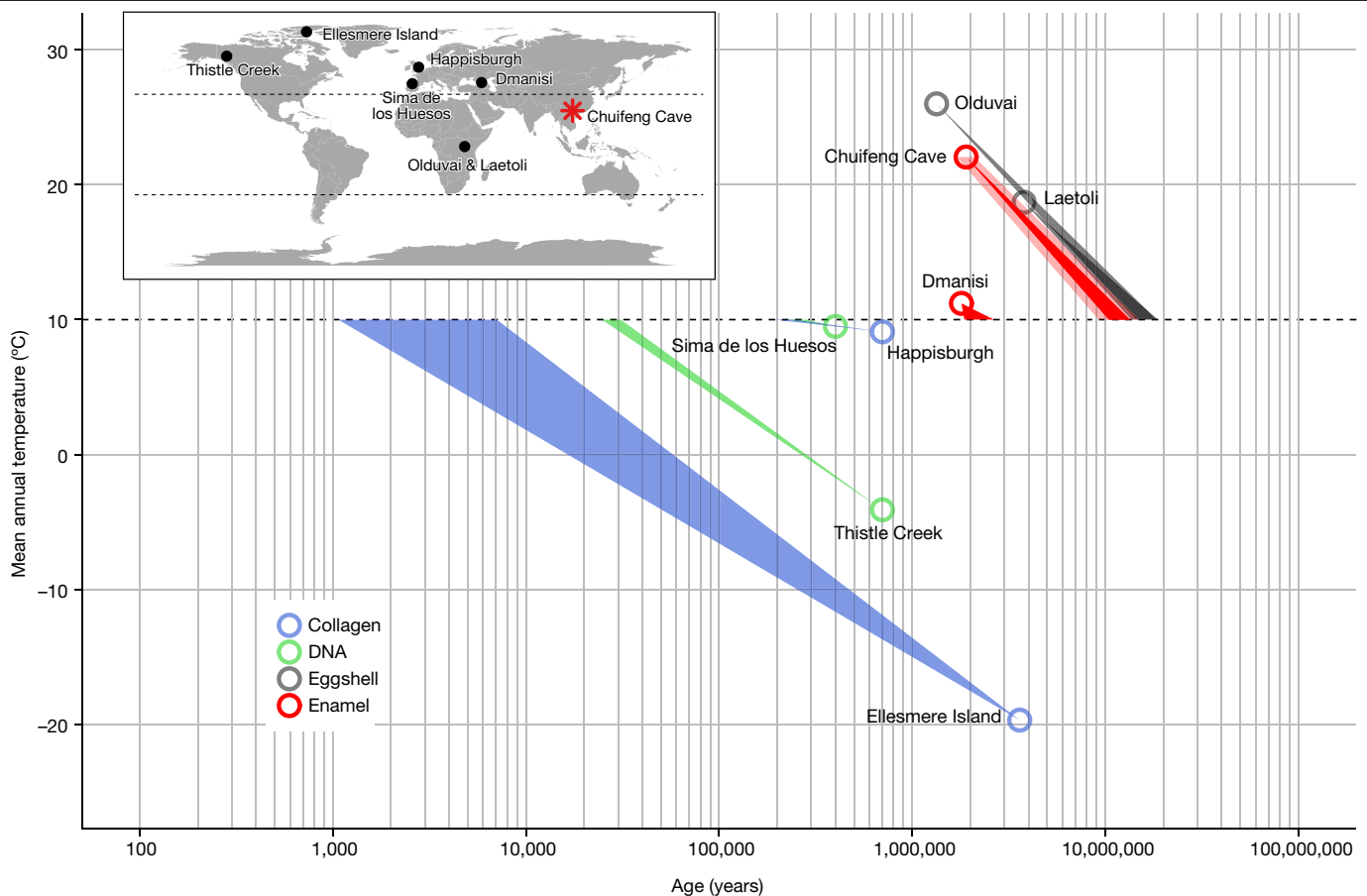


Fig. 1 | Thermal age of Chuifeng Cave, China, in the subtropics of Southeast Asia. Chronological (circles) and thermal (coloured wedges, @10 °C) ages of other Cenozoic ancient genomes and proteomes are given for comparison. For Chuifeng Cave, dark red shading indicates thermal age calculated at a chronological age of 1.9 Ma and lighter shading indicates 95% confidence

intervals based on a chronological age of 1.7 and 2.1 Ma, respectively. Inset, geographical location of Chuifeng Cave in the subtropics of Southeast Asia (red asterisk). Base map was generated using public domain data from <http://www.natureearthdata.com/>. The red asterisk also encloses the entire known geographical range of *G. blacki* fossils.

We identified no endogenous proteins from the dentine, but instead recovered an ancient enamel proteome composed of 409 unique peptides matching 6 endogenous proteins: amelogenin X (AMELX), ameloblastin (AMBN), amelotin (AMTN), enamelin (ENAM), matrix metalloproteinase-20 (MMP20) and alpha-2-HS-glycoprotein (AHSG, also known as FETUA; Extended Data Table 1). This observation extends the survival of ancient mammalian proteins to a thermal age, obtained by normalizing the chronological age to a constant temperature of 10 °C, of approximately 11.8 million years (Myr)@10 °C (Extended Data Table 2). Such a thermal age is well beyond the thermally oldest DNA (0.25 Myr@10 °C, Sima de los Huesos, Spain¹⁸), collagen (0.22 Myr@10 °C, Happisburgh, UK¹⁹) and enamel proteome (2.2 Myr@10 °C, Dmanisi, Georgia¹⁷) reported to date. The Chuifeng Cave enamel proteome is thus, to our knowledge, the oldest Cenozoic skeletal proteome currently reported (Fig. 1). The survival of a subtropical proteome from approximately 2 Ma suggests that chronologically older specimens from higher latitudes are likely to contain preserved ancient proteomes as well.

The content of the recovered enamel proteome is consistent with previously reported ancient enamel proteomes^{17,20,21}, with the addition of several peptides deriving from a single region of AHSG. Peptide matches to these proteins covered a minimum of 43 informative single amino acid polymorphisms (SAPs; Supplementary Table 3). In addition, the retrieved protein regions fell largely within areas previously recovered from an Early Pleistocene *Stephanorhinus* enamel proteome from Dmanisi¹⁷ (Supplementary Fig. 1). The absence of peptides specific to amelogenin Y (AMELY) suggests that the sampled molar might have

belonged to a female *Gigantopithecus*. Alternatively, male-diagnostic AMELY-specific peptides were not observed due to their degradation beyond the limit of detection of the instrument. The endogenous enamel proteome sequence coverage of 456 amino acids is lower than the previously recovered sequence coverage for a Dmanisi *Stephanorhinus* specimen (875 amino acids¹⁷; Supplementary Table 1). This observation is consistent with the older thermal age for Chuifeng Cave, compared to Dmanisi¹⁷.

We replicated enamel demineralization using two acids (TFA and HCl). The chromatograms of these two extracts showed that different peptide populations were released (Extended Data Fig. 3). Owing to the partial acidic hydrolysis²², which potentially occurs alongside demineralization, peptide populations with a wider range of acidity (Extended Data Fig. 4a) and hydrophobicity (Extended Data Fig. 4c) are generated using TFA. Our TFA-based demineralization returned 127 more unique non-overlapping peptide sequences than did the HCl-based demineralization (Extended Data Fig. 4e). The TFA extract, therefore, outperformed the HCl-based extraction, despite being carried out on a smaller amount of starting material¹⁷. Ultimately, the extended coverage of TFA-based demineralization increased the identification rate of informative SAPs, enhancing the phylogenetic information obtained (Extended Data Fig. 4d). Finally, the HCl- and TFA-demineralized samples had similar deamidation rates and average peptide lengths (Extended Data Fig. 5), which indicates that the two acids released peptide populations that were modified to the same extent.

The *Gigantopithecus* enamel proteome is characterized by extensive diagenetic modifications, such as high rates of deamidation (Fig. 2a) and

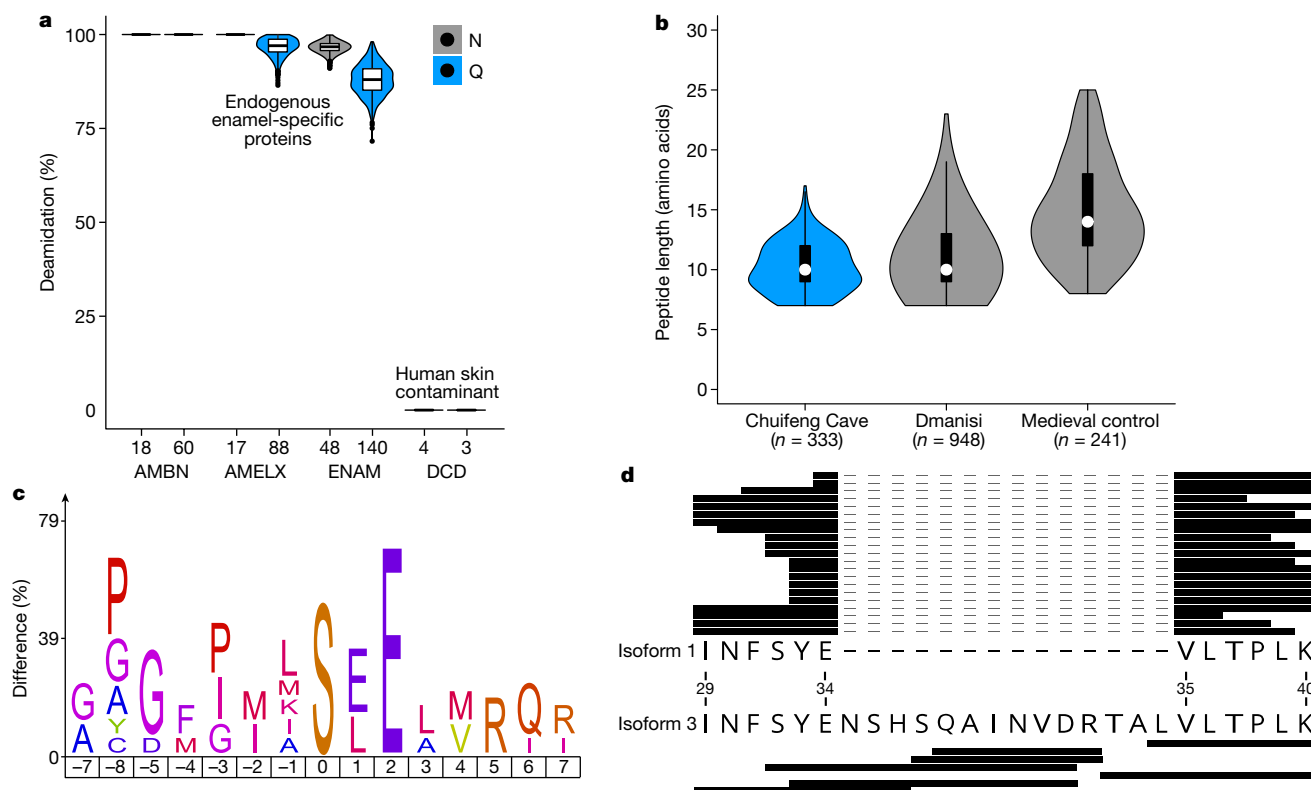


Fig. 2 | Modifications and degradation of *Gigantopithecus* enamel proteome. **a**, Violin plots of asparagine (N) and glutamine (Q) deamidation for selected proteins (n=1,000 bootstrap replicates of intensity-based peptide deamidation³²). Human dermcidin (DCD) is included as an example of a non-deamidated contaminant. For AMBN, all observed asparagines and glutamines are deamidated. For AMELX, all asparagines are deamidated. For DCD, no observed asparagines and glutamines are deamidated. The number of peptides used for the calculation are shown at the bottom. **b**, Violin plots of peptide lengths for *Gigantopithecus*, an Early Pleistocene rhinoceros from Dmanisi and a Medieval control sample¹⁷. **c**, Sequence-motif analysis of the

over-representation of specific amino acids around the phosphorylated amino acid (position 0; n=14). **d**, Peptide coverage of AMELX protein isoforms. Matching peptides are indicated by black bars for isoform 1 (n=21) and isoform 3 (n=7). The latter includes an insertion due to alternative splicing between isoform 1, amino acid positions 34 and 35 (coordinates in reference to UniProt Accession number: Q99217-1 [AMELX_HUMAN]). **b** includes data on AMELX, AMBN, ENAM, AMTN and MMP20 only. For **a** and **b**, box plots define the range of the data (whiskers extending to 1.5× the interquartile range), 25th and 75th percentiles (boxes) and medians (dots/lines). For **a**, outliers are indicated as black dots (beyond 1.5× the interquartile range).

a high degree of degradation, as indicated by relatively short peptide lengths (Fig. 2b); this is expected for an ancient proteome preserved in subtropical conditions. When quantifying peptide intensities using label-free quantification (LFQ) implemented in MaxQuant²³, we observed that summed and normalized MS1 spectral intensities were higher for shorter peptides than for longer peptides (Extended Data Fig. 4b). Finally, the peptide lengths of the Chuifeng Cave enamel proteome were shorter than those identified in thermally younger enamel proteomes (Fig. 2b).

Enamel-specific proteins are modified in vivo through protein phosphorylation, alternative splicing of AMELX, and MMP20- and KLK4-mediated proteolysis. Such modifications could survive in ancient proteomes. We detected evidence of surviving in vivo post-translational modifications, such as serine phosphorylation in the S-x-E/pHS motif, which is recognized by the secreted kinase FAM20C (Fig. 2c). The FAM20C kinase regulates the phosphorylation of extracellular proteins involved in biomineralization²⁴. We also observed two alternative-splicing-derived isoforms of AMELX (Fig. 2d). These observations are similar to those from other Early Pleistocene enamel proteomes¹⁷. The *Gigantopithecus* enamel proteome therefore demonstrates that such in vivo modifications can be recovered from hominid samples across the Pleistocene.

To achieve a protein-based phylogenetic placement of *Gigantopithecus*, we compared the enamel proteome sequences we retrieved with those of extant apes (Hominoidea). We used publicly available whole-genome sequence data to predict enamel protein sequences from relevant species^{25,26} (Supplementary Table 2, Supplementary Figs. 2–12). Our results show that *Gigantopithecus* represents a sister taxon to all extant

orangutans (*Pongo* spp.) and forms a monophyletic group with extant pongines (Fig. 3a, Extended Data Figs. 6, 7). We then attempted to estimate the divergence time between *Gigantopithecus* and *Pongo* species using two approaches: (1) a pairwise distance approach and (2) a Bayesian approach using MrBayes (see Methods). Although confidence intervals obtained for the divergence estimates of the *Pongo–Gigantopithecus* split are large, our results indicate that *Gigantopithecus* diverged from the extant *Pongo* species in the Middle or Late Miocene (approximately 10 Ma or 12 Ma using the Bayesian or pairwise distance approaches, respectively; Fig. 3b). This suggests that, despite an exclusively Pleistocene fossil record, *Gigantopithecus* is a member of an early radiation of pongines, whose diversity peaked during the Middle and Late Miocene (Fig. 3b). Our results thereby resolve not only the phylogenetic position of *Gigantopithecus*, but also renew the debate on the evolutionary relationships between extant hominids and early hominids present in the fossil record².

The presence of AHSG in the *Gigantopithecus* proteome is intriguing, as this protein is not commonly observed in (modern) hominid enamel proteomes. All retrieved peptides derive from a single, highly conserved region that is bordered by disulfide cysteine bonds on either side (Extended Data Fig. 8). AHSG is highly glycosylated in vivo, but we observed no glycosylation during our bioinformatics analysis. The observed sequence contains regularly spaced aspartic acid residues that provide a suitable motif for binding to basic calcium phosphate lattices²⁷. The notion that this specific peptide sequence is involved in biomineral binding is supported by three observations: (1) this region is presented on the external surface of AHSG²⁸; (2) such surfaces have

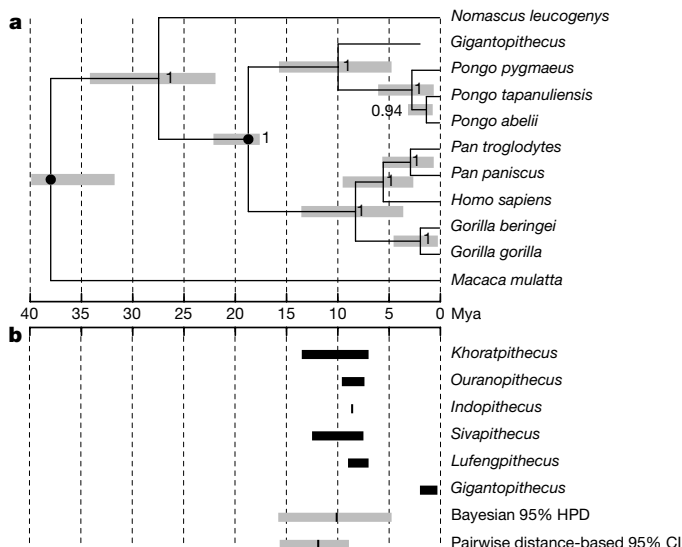


Fig. 3 | Bayesian phylogeny and divergence of *Gigantopithecus* among extant apes (Hominoidea). **a**, Time-calibrated Bayesian phylogeny of *Gigantopithecus*. Circled nodes were fixed for topology (see Methods). Grey error bars represent the 95% highest posterior density (HPD) intervals for the divergence estimates. **b**, Distribution of probable and possible extinct pongines known from the fossil record, including *Gigantopithecus* (black bars). Grey bars represent the 95% HPD interval obtained from the Bayesian approach and the 95% confidence interval obtained from the pairwise-distance-based approach, of the *Gigantopithecus*-*Pongo* divergence.

been demonstrated to bind biominerals in other systems as well²⁹; and (3) this type of binding enhances peptide preservation²⁹. AHSG acts as a key component of bone and dentine mineralization by inhibiting the extrafibrillar mineralization of collagen type I helices³⁰ and has previously been hypothesized to have a role in amelogenesis⁹. Our extracts contained no endogenous plasma proteins, such as human serum albumin, or other common dentine proteins, such as collagen type I. We also did not identify any AHSG peptides in our dentine sample. We therefore exclude the possibility that the AHSG peptides derived from dentine. *Gigantopithecus* is known to have relatively long enamel formation times and thick enamel compared to several extant and extinct hominids, including its phylogenetically closest relatives^{10,31}. We therefore hypothesize that *Gigantopithecus* recruited AHSG as an additional molecular component to favour enamel biomineralization during prolonged amelogenesis, ultimately playing a part comparable to the one it has in bone and dentine mineralization⁹.

Our results reveal the long-debated phylogenetic position of *Gigantopithecus* as an early diverging pongine. We demonstrate that ancient enamel proteomes can be retrieved from Early Pleistocene samples preserved in subtropical conditions, well beyond the current limitations of biomolecular research in hominid and hominin evolution. In addition, the survival of an Early Pleistocene *Gigantopithecus* enamel proteome enables us to assess the presence of multiple forms of in vivo modification. Finally, palaeoproteomic analysis allowed us to identify a hitherto unknown biological component of tooth formation in an extinct hominid. These findings suggest that the palaeoproteomic analysis of hominid enamel has the potential to provide a molecular perspective on human and great ape evolution.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1728-8>.

- Zhang, Y. & Harrison, T. *Gigantopithecus blacki*: a giant ape from the Pleistocene of Asia revisited. *Am. J. Phys. Anthropol.* **162** (Suppl. 63), 153–177 (2017).
- Harrison, T. Apes among the tangled branches of human origins. *Science* **327**, 532–534 (2010).
- Begun, D. R. How to identify (as opposed to define) a homoplasy: examples from fossil and living great apes. *J. Hum. Evol.* **52**, 559–572 (2007).
- Kelley, J. In *The Primate Fossil Record* (ed. Hartwig, W. C.) 369–384 (Cambridge Univ. Press, 2002).
- Miller, S. F., White, J. L. & Ciochon, R. L. Assessing mandibular shape variation within *Gigantopithecus* using a geometric morphometric approach. *Am. J. Phys. Anthropol.* **137**, 201–212 (2008).
- Grehn, J. R. & Schwartz, J. H. Evolution of the second orangutan: phylogeny and biogeography of hominid origins. *J. Biogeogr.* **36**, 1823–1844 (2009).
- Shao, Q. et al. ESR, U-series and paleomagnetic dating of *Gigantopithecus* fauna from Chufeng Cave, Guangxi, southern China. *Quat. Res.* **82**, 270–280 (2014).
- Wang, W. New discoveries of *Gigantopithecus blacki* teeth from Chufeng Cave in the Bubing Basin, Guangxi, south China. *J. Hum. Evol.* **57**, 229–240 (2009).
- Bartlett, J. D. et al. Protein-protein interactions of the developing enamel matrix. *Curr. Top. Dev. Biol.* **74**, 57–115 (2006).
- Dean, M. C. & Schrenk, F. Enamel thickness and development in a third permanent molar of *Gigantopithecus blacki*. *J. Hum. Evol.* **45**, 381–388 (2003).
- Von Koenigswald, G. H. R. Eine fossile Säugetierfauna mit Simia aus Sudchina. *Proc. K. Ned. Akad. Wet.* **38**, 872–879 (1935).
- Zhang, Y. et al. New 400–320 ka *Gigantopithecus blacki* remains from Hejiang Cave, Chongzuo City, Guangxi, South China. *Quat. Int.* **354**, 35–45 (2014).
- Zhao, L. X. & Zhang, L. Z. New fossil evidence and diet analysis of *Gigantopithecus blacki* and its distribution and extinction in South China. *Quat. Int.* **286**, 69–74 (2013).
- Pei, W. C. Excavation of Liucheng *Gigantopithecus* cave and exploration of other caves in Kwangsi. *Memoir of the Institute of Vertebrate Palaeontology and Palaeoanthropology, Academia Sinica* **7**, 1–54 (1965).
- Bocherens, H. et al. Flexibility of diet and habitat in Pleistocene South Asian mammals: implications for the fate of the giant fossil ape *Gigantopithecus*. *Quat. Int.* **434**, 148–155 (2017).
- Ciochon, R. et al. Dated co-occurrence of *Homo erectus* and *Gigantopithecus* from Tham Khuyen Cave, Vietnam. *Proc. Natl Acad. Sci. USA* **93**, 3016–3020 (1996).
- Cappellini, E. et al. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature* **574**, 103–107 (2019).
- Meyer, M. et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).
- Wadsworth, C. & Buckley, M. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* **28**, 605–615 (2014).
- Stewart, N. A. et al. The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction. *RSC Advances* **6**, 61673–61679 (2016).
- Castiblanco, G. A. et al. Identification of proteins from human permanent erupted enamel. *Eur. J. Oral Sci.* **123**, 390–395 (2015).
- Cristobal, A. et al. Toward an optimized workflow for middle-down proteomics. *Anal. Chem.* **89**, 3318–3325 (2017).
- Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
- Tagliabracci, V. S. et al. Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* **336**, 1150–1153 (2012).
- Prado-Martinez, J. et al. Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
- Nater, A. et al. Morphometric, behavioral, and genomic evidence for a new orangutan species. *Curr. Biol.* **27**, 3487–3498 (2017).
- Tang, N. & Skibsted, L. H. Calcium binding to amino acids and small glycine peptides in aqueous solution: toward peptide design for better calcium bioavailability. *J. Agric. Food Chem.* **64**, 4376–4389 (2016).
- Heiss, A. et al. Structural basis of calcification inhibition by alpha 2-HS glycoprotein/fetuin-A. Formation of colloidal calciprotein particles. *J. Biol. Chem.* **278**, 13333–13341 (2003).
- Demarchi, B. et al. Protein sequences bound to mineral surfaces persist into deep time. *eLife* **5**, e17092 (2016).
- Price, P. A., Torian, D. & Lim, J. E. Mineralization by inhibitor exclusion: the calcification of collagen with fetuin. *J. Biol. Chem.* **284**, 17092–17101 (2009).
- Kono, R. T., Zhang, Y., Jin, C., Takai, M. & Suwa, G. A 3-dimensional assessment of molar enamel thickness and distribution pattern in *Gigantopithecus blacki*. *Quat. Int.* **354**, 46–51 (2014).
- Mackie, M. et al. Palaeoproteomic profiling of conservation layers on a 14th century Italian wall painting. *Angew. Chem. Int. Ed.* **57**, 7369–7374 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Chuifeng Cave

The Chuifeng Cave (23° 34' 27" N, 107° 00' 22" E) is one of the most representative sites for the Early Pleistocene *G. blacki* fauna⁸. The site is located in the Bubing Basin in the northwestern part of the Guangxi Zhuang Autonomous Region, south China (Extended Data Fig. 1). The cave is 19 m long, 0.5–2 m wide and 1.5–5 m high, penetrating the limestone from southeast to northwest at a height of ~77 m above the local valley floor. A fossiliferous sandy clay with a few limestone breccias fills most of the cave, with an average depth of 1.3 m (Extended Data Fig. 2). Four excavation areas (A, B, C and D) were excavated down to limestone bedrock at 10-cm intervals. Twenty-four large mammalian species, including 92 *G. blacki* teeth, were unearthed from the cave⁸. The Chuifeng Cave mammalian fauna is characterized by the occurrence of typical Early Pleistocene species, such as *Hystrix magna*, *Sinomastodon* sp., *Stegodon preorientalis*, *Ailuropoda microta*, *Pachycrocuta licenti*, *Tapirus sanyuanensis* and *Sus peii*⁸. This mammalian fauna is comparable with other *Gigantopithecus*-containing faunas of the Early Pleistocene in southern China, such as Baikong³³, Longgupo³⁴ and Liucheng³⁵. The mammalian fauna composition is consistent with the results of combined electron spin resonance (ESR)/U-series dating and sediment palaeomagnetic studies (~1.9 Ma)³⁶. In the present study, we collected one well-preserved *G. blacki* tooth (excavation number CF-B-16) for palaeoproteomic analysis. This tooth was excavated from area B at a depth of 90 cm from the sediment surface and, based on its stratigraphic position, dated to ~1.9 Ma. No other specimens were tested before CF-B-16, and no specific selection was made as to which *Gigantopithecus* tooth would be analysed.

Thermal age

Thermal age was calculated to allow comparison with previously published ancient genomes, ancient proteomes and collagen peptide mass fingerprinting studies, from other temporal and geographical localities. Temperature estimates for the hominin occupation of Dmanisi based on herpetological fauna suggest a temperature about 3.1 °C above the current MAT³⁷, while the sea surface temperature record used²⁹ predicts a negative ΔT at the time of hominin occupation. Given this discrepancy and the widely different temperature estimates for the Last Glacial Maximum (LGM) in the Caucasus, we conservatively use a scale factor of 0, correlating with a ΔT of approximately -0.2 °C and a current MAT of 11.2 °C. Our thermal age prediction for Dmanisi (2.2 Myr@10 °C) should therefore be seen as conservative. The thermal age for Chuifeng Cave was calculated with a general lapse rate between mean annual temperature (MAT) and altitude of 5.0 °C/km, a scale factor of 0.7 and a ΔT at LGM of -3 °C. The actual ΔT at LGM might have been more pronounced, again leading to a conservative estimate for the thermal age of Chuifeng Cave also. MAT was estimated based on the ten closest weather stations listed in publicly accessible World Meteorological Organization (WMO) data (Extended Data Table 2). Thermal age calculations are, among other factors, altitude dependent, but only five out of these ten weather stations have altitudes directly associated with them. We therefore estimated the altitude of the other five weather stations through an online resource (<https://www.advancedconverter.com/map-tools/find-altitude-by-coordinates>). The correlation between WMO altitude and estimated altitude was $R^2 = 0.99$, providing sufficient validity to our estimated altitudes. The MATs for all weather stations were then averaged to obtain an approximate MAT for Chuifeng Cave. Next, thermal age was calculated for chronological ages of 1.7 Myr, 1.9 Myr and 2.1 Myr, giving estimates of the minimum (9.2 Myr@10 °C), maximum (15.0 Myr@10 °C) and mean (11.8 Myr@10 °C) thermal ages associated with the Chuifeng Cave fauna within a 95% confidence interval (Fig. 1). The Chuifeng Cave proteome is therefore, to our knowledge, substantially older than the oldest collagen peptide mass fingerprint (Ellesmere Island, 0.003 Myr@10 °C), oldest mammalian genome

(Thistle Creek, 0.03 Myr@10 °C), oldest hominin genome (Sima de los Huesos, 0.25 Myr@10 °C) and oldest enamel proteome (Dmanisi, 2.2 Myr@10 °C) published to date²⁹. Full thermal age calculations can be found in Supplementary Data 2.

Protein extraction

Extraction of ancient proteins took place in facilities at the Natural History Museum of Denmark dedicated to extracting ancient DNA and ancient proteins. These laboratories include clean rooms fitted with filtered ventilation and positive air pressure³⁸. A negative extraction blank was processed alongside the ancient extractions, with the additional inclusion of injection blanks during tandem mass spectrometry (MS/MS) analysis to monitor potential protein contamination during all stages of analysis.

Two samples of enamel (185 and 118 mg) and one of dentine (192 mg) were removed from the same molar (CF-B-16) using a sterilized drill and crushed to a rough powder. One enamel sample and the dentine sample were demineralized in 1.2 M HCl at 3 °C for 24 h, and the other enamel sample was demineralized at the same temperature and duration using 10% TFA. Subsequently, solubilized protein residues were cleaned, concentrated and immobilized on C18 Stage-Tips using previously published methods¹⁷. No other samples from Chuifeng Cave were analysed before or during the analysis of CF-B-16.

LC-MS/MS analysis

The extracts were analysed by nanoflow liquid chromatography–tandem mass spectrometry (nanoLC-MS/MS) using a 15-cm capillary column (75 µm inner diameter, packed with 1.9 µm C18 beads (Reprosil-AQ Pur, Dr. Maisch)) on an EASY-nLC 1200 system (Proxeon) connected to a Q-Exactive HF-X mass spectrometer (Thermo Scientific). The nanoLC gradient and MS parameters followed a previously published Q-Exactive HF-X method³². System wash blanks were performed before and after every sample to hinder cross-contamination.

Database construction

We constructed a protein sequence database for Hominoidea proteins known to be present in enamel proteomes (Supplementary Table 2), to which we added the homologous sequences from one Cercopithecoid (*Macaca mulatta*) as an outgroup for phylogenetic analysis. As few protein sequences are publicly available for *Pongo pygmaeus*, we predicted those sequences from publicly available genomic sequence data using the known gene coordinates of *Pongo abelii* homologues. Similarly, we generated de novo AMELY sequences for *P. abelii* and *P. pygmaeus*. Finally, we added common laboratory contaminants to allow spectra from such proteins to be confidently identified³⁸.

Ancestral sequence reconstruction. Previous research indicates that cross-species proteomic effects, observed during spectral identification, substantially reduce the identification of phylogenetically informative amino acid positions at large evolutionary distances³⁹. We reasoned that this was likely to occur in the case of *Gigantopithecus* proteins⁴⁰, and therefore reconstructed the ancestral protein sequences of enamel-specific proteins. Ancestral sequence reconstruction (ASR) was conducted across the entire Hominoidea phylogeny using PhyloBot⁴¹. Input sequences were constrained phylogenetically to (*Macaca*, (*Nomascus*, ((*Pongo abelii*, *Pongo pygmaeus*), *Gorilla*), (*Homo*, ((*Pan paniscus*, *Pan troglodytes*))))). We added those sequences to the reference protein database to account for them in the database search of PEAKS and MaxQuant.

Isoform variation. After obtaining complete protein sequences for all extant hominids, we added isoforms not present in UniProt or GenBank for the proteins AMELX, AMELY, AMBN, AMTN, KLK4 and TUFT1, including the reconstructed ASR sequences of these proteins, to the database. We assumed that the isoforms for these non-human hominids would

result from identically placed alternative splicing across species and ancestral nodes (as also supported by all UniProt isoforms present for the studied proteins). Thus, we copied these alternative splicing sites onto the available reference sequences to create the missing isoforms. Database sequence names for these proteins were appended with ‘_ManIso2’ or ‘_ManIso3’.

Proteomic data analysis

Raw mass spectrometry data were searched per sample type (enamel, dentine, extraction blank and injection blanks) against a sequence database containing all common enamel proteins for all extant hominids (see above). We used PEAKS⁴² (v.7.5) and MaxQuant²³ (v.1.6.2.6) software. The de novo and error-tolerant implementations of PEAKS, and the dependent peptide algorithm implemented in MaxQuant, were used to generate possible, additional, SAP variation in enamel protein sequences. Such novel SAPs could represent unique amino acid substitutions on the *Gigantopithecus* lineage, which are not relevant to its phylogenetic placement but are relevant for dating the *Pongo*–*Gigantopithecus* divergence. Next, these potential sequence variants were added to a newly constructed sequence database and verified in separate searches in PEAKS and MaxQuant. We defined as variable modifications methionine oxidation, proline hydroxylation, glutamine and asparagine deamidation, pyro-glutamic acid from glutamic acid, pyro-glutamic acid from glutamine and phosphorylation (STY). No fixed modifications were selected. We did not use an enzymatic protease during sample preparation, therefore the digestion mode was set to ‘unspecific’. For PEAKS, peptide spectrum matches were only accepted with a false-discovery rate (FDR) $\leq 1.0\%$, and precursor mass tolerance was set to 10 ppm and fragment mass tolerance to 0.05 Da. For MaxQuant, peptide spectrum matches and protein FDR were set at $\leq 1.0\%$, with a minimum Andromeda score of 40 for all peptides. Protein matches were accepted with a minimum of two unique peptide sequences in at least one of the MaxQuant or PEAKS searches, including the removal of non-specific peptides after BLASTp searches of peptides matching non-enamel proteins against UniProt and GenBank databases. Proteins that are retained after applying these criteria are listed in Extended Data Table 1. Examples of annotated MS/MS spectra after MaxQuant analysis can be found in Supplementary Figs. 3–12.

Assessment of protein damage and degradation followed protocols explained elsewhere^{17,32,43} and included rates of deamidation and a comparison of observed peptide lengths. GRAVY index scores of peptide hydrophobicity were calculated using the R package Peptides, with the scale set to ‘KyteDoolittle’.

Phylogenetic and divergence analysis

Comparative reference dataset. We assembled a reference dataset with five protein sequences retrieved from the ancient sample (AMBN, AMELX, AMTN, ENAM and MMP20) and relevant extant species (Supplementary Table 2). Protein sequences for human (*Homo sapiens*), common chimpanzee (*P. troglodytes*), bonobo (*P. paniscus*), Sumatran orangutan (*P. abelii*), Western gorilla (*Gorilla gorilla*), rhesus macaque (*M. mulatta*) and the white-cheeked gibbon (*Nomascus leucogenys*) were obtained from the UniProt database. Additionally, we expanded our dataset with protein sequences from publicly available whole-genome sequence data from present-day great apes (in total 27 orangutans, 42 gorillas, 11 bonobos and 61 chimpanzees^{25,26,44}), as well as 19 human individuals from the Simons Genome Diversity Project (SGDP)⁴⁵. See Supplementary Information for the human sample numbers taken from the SGDP dataset.

Reconstruction of protein sequences from whole-genome sequencing data. DNA sequence reads for reference samples used were mapped to the human genome (version hg19) using BWA-MEM v.0.7.5a-r405 (<http://bio-bwa.sourceforge.net/bwa.shtml>) with default

parameters. PCR and optical duplicates were identified and removed using PICARD v.1.91 (<https://sourceforge.net/projects/picard/files/picard-tools/1.91/>). Single-nucleotide polymorphisms were called on the read alignments using the GATK UnifiedGenotyper (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_genotyper_UnifiedGenotyper.php).

To reconstruct the protein sequences from the genotype calls, we first created a consensus sequence for each of the five genes of interest and for each sample. Indels were not considered and a random allele was chosen at heterozygous positions. Next, we removed the intron sequences from each gene using the annotation of the reference human genome (hg19) available in the ENSEMBL database. For each of the in silico spliced genes, we performed a tblastn search⁴⁶ using the human reference protein as the query. Finally, we obtained the translated protein sequences from the resulting alignments.

Assessing the phylogenetic position of *G. blacki*. We compared the *G. blacki* protein sequences with the corresponding homologues of the species in the reference panel. For each gene, we built two multiple sequence alignments using mafft⁴⁷. The first incorporated all samples in the reference panel ($n = 164$). The second incorporated only a single sample per species (Supplementary Table 3). To account for isobaric amino acids (leucines and isoleucines), which cannot be distinguished in the ancient protein data, we changed all isoleucines to leucines at positions where the ancient sample carried either of those amino acids. To assess the phylogenetic position of the ancient sample, two inference approaches were used: a maximum-likelihood and a Bayesian inference.

Maximum-likelihood approach. PhyML v.3.1⁴⁸ was used to infer a maximum-likelihood tree, branch lengths and substitutions rates for each individual protein alignment (Supplementary Fig. 2) and for the concatenated alignment. For each alignment, we started from three random trees (--n_rand_starts 3 -s BEST --rand_start), used the JTT model (-m JTT -f m), and obtained maximum likelihood estimates for the gamma distribution shape parameter (-a e) and the proportion of invariable sites (-v e). Support values were obtained for each bipartition based on 100 non-parametric bootstrap replicates. The bootstrap results per branch split are shown in Extended Data Fig. 6b.

Bayesian approach. As a complementary approach, we used MrBayes⁴⁹ and the concatenated alignment to infer the phylogenetic position of the ancient sample (Fig. 3, Extended Data Fig. 6b). We set an independent bipartition for each gene and estimated: substitution rates, across-site rate variation and the proportion of invariable sites (unlink Statefreq = (all) Ratemultiplier = (all) Aamodel = (all) Shape = (all) Pinvar = (all)). MrBayes was executed using the CIPRES portal⁵⁰. The MCMC algorithm was set to 5,000,000 cycles with 4 chains and a temperature parameter of 0.2. The convergence of the algorithm was assessed using Tracer v.1.6.0 after discarding 25% of the iterations as burn-in. MrBayes was run against the reference sequence for each species (Extended Data Fig. 6c) or against 162 great ape individuals, one hylobatid and one cercopithecoid (Extended Data Fig. 7). Both of these analyses, as well as the PHyML maximum likelihood approach, resulted in the same topology. The analysis using a large number of individuals shows, however, that resolution within the genus *Pongo* is limited (Extended Data Fig. 7). Nevertheless, the placement of *Gigantopithecus* is fully supported.

Divergence time of *Gigantopithecus*. We estimated the divergence time between *Gigantopithecus* and the *Pongo* branch first by using a distance-based approach. We used the alignment of the amino acid sequences of reference genome sequences for each species as well as diversity data (see above). A distance matrix was created from the concatenated protein sequences of all individuals using the function dist.ml from the R package phangorn⁵¹ under the LG amino acid substitution model⁵². We used pairwise exclusion to increase the amount of data for the present-day branches. We then calculated the mean

Article

difference of all orangutan sequences to all sequences from *Homo*, *Pan* and *Gorilla*, and the mean difference of all orangutan sequences to *Gigantopithecus* (Extended Data Fig. 6a). We used the average distance between orangutan and the other extant great apes as a scaling factor, assuming a divergence time between these branches⁵³ of 23.8 Ma. Under this assumption, the molecular divergence of *Gigantopithecus* from the *Pongo* branch is 9.98 Ma. However, because Chuifeng Cave is dated to 1.9 Ma, this branch is likely to be underestimated and its age needs to be corrected to 11.88 Ma. We combine the 95% confidence interval of the distance matrix with the 95% confidence interval of the mutation rate estimate⁵³, and add the upper and lower values of the 95% confidence interval for the Chuifeng Cave dating (1.7–2.1 Ma), and therefore suggest conservative upper and lower boundaries for the divergence of 8.91 and 15.65 Ma, respectively.

If mutation rates did not substantially differ between extant *Pongo* and *Gigantopithecus*, this estimate should reflect the molecular evolution of their common branch. We calculated the divergence between the other great apes, taking into account the mutation rate differences on these lineages as scaling factors⁵³. The resulting divergence time between *Gorilla* and the *Homo/Pan* branch is estimated at 10.27 Ma (7.9–13.25 Ma, 95% confidence interval), and the divergence between *Homo* and *Pan* at 8.72 Ma (8.06–13.81 Ma, 95% confidence interval). These values are in strong agreement with previous estimates⁵³, suggesting that these protein sequences represent well the known phylogeny of the great apes. Clearly, all divergence time estimates scale with assumptions of mutation rates. We also caution that the small number of mutations in the peptide fragments in *Gigantopithecus* constitutes a severe limitation on the precision of these estimates on this branch. However, the phylogenetic position of *Gigantopithecus* as a sister clade to orangutans is also well supported in this analysis: a phylogenetic tree from a distance matrix of the reference sequences for these species (neighbour joining tree in phangorn; maximum likelihood computed with the pml function; 1,000 bootstrap replicates) separates *Gigantopithecus* from orangutans with 100% bootstrap support.

We used the program MrBayes⁴⁹ to estimate divergence time estimates in a Bayesian framework using the reference genome sequences. We defined *M. mulatta* as outgroup, grouped *Pan*, *Homo* and *Gorilla* together as well as *Pongo* and *Gigantopithecus*, and set the divergence time of the two groups with a uniform distribution of 17.739–26.061 Ma, using a previously published estimate⁵³. Furthermore, we set the divergence time of the macaques and apes at 26.061–39.9 Ma (from the maximum divergence time of the hominids to a very high divergence time of the apes). We used a variable mutation rate and the VT amino acid substitution model⁵⁴ in five million iterations. This results in a divergence time for *Gigantopithecus*–*Pongo* of 10.14 Ma (4.76–15.79 Ma, 95% HPD interval). The divergence of *Gorilla* from the *Homo/Pan* branch is estimated at 8.59 Ma (4.62–13.56 Ma, 95% HPD interval), and the divergence of *Homo* and *Pan* at 5.78 Ma (2.64–9.53 Ma, 95% HPD interval). These are largely consistent with, but slightly younger than, previous estimates^{53,55}, possibly owing to a mutation slowdown in these lineages compared to the *Pongo* lineage, which is not taken into account here. However, they seem in agreement with the fossil record indicating the origin of hominins around 6–8 Ma and the dating of a possible early Gorillini (*Chororapithecus*) around 7–9 Ma^{55–59}. Therefore, we conclude that the relative branch lengths of the tree (Fig. 3b) are concordant with the overall phylogeny and the estimates presented above.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All the mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD013838. Generated ancient protein consensus sequences for *G. blacki* can be found in Supplementary Data 1. Full thermal age calculations can be found in Supplementary Data 2.

33. Jin, C. et al. Chronological sequence of the early Pleistocene *Gigantopithecus* faunas from cave sites in the Chongzuo, Zuojiang River area, South China. *Quat. Int.* **354**, 4–14 (2014).
34. Huang, W. C. R. et al. Early *Homo* and associated artefacts from Asia. *Nature* **378**, 275–278 (1995).
35. Pei, W. Discovery of *Gigantopithecus* mandibles and other material in Liucheng district of central Kwangsi in South China. *Vertebrata Palasiatica* **1**, 65–71 (1957).
36. Sun, L. et al. Magnetochronological sequence of the Early Pleistocene *Gigantopithecus* faunas in Chongzuo, Guangxi, southern China. *Quat. Int.* **354**, 15–23 (2014).
37. Blain, H.-A., Agustí, J., Lordkipanidze, D., Rook, L. & Delfino, M. Paleoclimatic and paleoenvironmental context of the Early Pleistocene hominins from Dmanisi (Georgia, Lesser Caucasus) inferred from the herpetofaunal assemblage. *Quat. Sci. Rev.* **105**, 136–150 (2014).
38. Hendy, J. et al. A guide to ancient protein studies. *Nat. Ecol. Evol.* **2**, 791–799 (2018).
39. Welker, F. Elucidation of cross-species proteomic effects in human and hominin bone proteome identification through a bioinformatics experiment. *BMC Biol.* **18**, 23 (2018).
40. Welker, F. Palaeoproteomics for human evolution studies. *Quat. Sci. Rev.* **190**, 137–147 (2018).
41. Hanson-Smith, V. & Johnson, A. PhyloBot: a web portal for automated phylogenetics, ancestral sequence reconstruction, and exploration of mutational trajectories. *PLoS Comput. Biol.* **12**, e1004976 (2016).
42. Zhang, J. et al. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012).
43. Welker, F. et al. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl Acad. Sci. USA* **113**, 11162–11167 (2016).
44. de Manuel, M. et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**, 477–481 (2016).
45. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
46. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
47. Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144–3146 (2012).
48. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
49. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
50. Miller, M. A., Pfeiffer, W. & Schwartz, T. In *Gateway Computing Environments Workshop (GCE)* 1–8 (New Orleans, 2010).
51. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
52. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
53. Besenbacher, S., Hvilsom, C., Marques-Bonet, T., Mailund, T. & Schierup, M. H. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat. Ecol. Evol.* **3**, 286–292 (2019).
54. Müller, T. & Vingron, M. Modeling amino acid replacement. *J. Comput. Biol.* **7**, 761–776 (2000).
55. Langergraber, K. E. et al. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl Acad. Sci. USA* **109**, 15716–15721 (2012).
56. Katoh, S. et al. New geological and palaeontological age constraint for the gorilla–human lineage split. *Nature* **530**, 215–218 (2016).
57. Senut, B. et al. First hominid from the Miocene (Lukeino Formation, Kenya). *Comptes Rendus Acad. Sci. IIA Earth Planet. Sci.* **332**, 137–144 (2001).
58. Brunet, M. et al. A new hominid from the Upper Miocene of Chad, central Africa. *Nature* **418**, 145–151 (2002).
59. Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).

Acknowledgements E.C. and F.W. are supported by VILLUM FONDEN (17649) and by the European Commission through a Marie Skłodowska-Curie (MSCA) Individual Fellowship (795569). T.M.-B. is supported by BFU2017-86471-P (MINECO/FEDER, UE), NIHM grant U01 MH106874, Howard Hughes International Early Career grant, Obra Social ‘La Caixa’ and Secretaria d’Universitats i Recerca and CERCA Programme del Departament d’Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880). E.C., J.C., J.V.O., D.S. and P.G. are supported by the Marie Skłodowska-Curie European Training Network (ETN) TEMPRA, a project funded by the European Union’s EU Framework Program for Research and Innovation Horizon 2020 under grant agreement no. 722606. M.J.C. and M.M. are supported by the Danish National Research Foundation award PROTEIOS (DNRF128). Work at the Novo Nordisk Foundation Center for Protein Research is funded in part by a donation from the Novo Nordisk Foundation (NNF14CC0001). Research at Chuifeng Cave is made possible by support from the National Natural Science Foundation of China (41572023) and by a grant from the Bagui Scholar of Guangxi. M.K. was supported by a Deutsche Forschungsgemeinschaft (DFG) fellowship (KU 3467/1-1) and the Postdoctoral Junior Leader Fellowship Programme from ‘la Caixa’ Banking Foundation (LCF/BQ/PR19/11700002). M.E.A. is supported by the Independent Research Fund Denmark (7027-00147B). We thank E. Willerslev for critical reading of the manuscript, scientific support and guidance.

Author contributions F.W., E.C., F.D. and T.M.-B. designed the study. W.W. conducted excavation of Chuifeng Cave. W.W. and W.L. carried out faunal analysis of Chuifeng Cave. W.W., W.L. and M.E.A. provided ancient samples. F.W., J.R.-M., M.K., P.G., D.S., M.M., J.V.O. and M.d.M. performed data generation and analysed data with support from A.-M.B., M.J.C., J.C., C.L.-F., F.D. and T.M.-B. F.W., E.C. and T.M.-B. wrote the manuscript with contributions from all authors.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1728-8>.

Correspondence and requests for materials should be addressed to F.W., W.W., T.M.-B. or E.C.

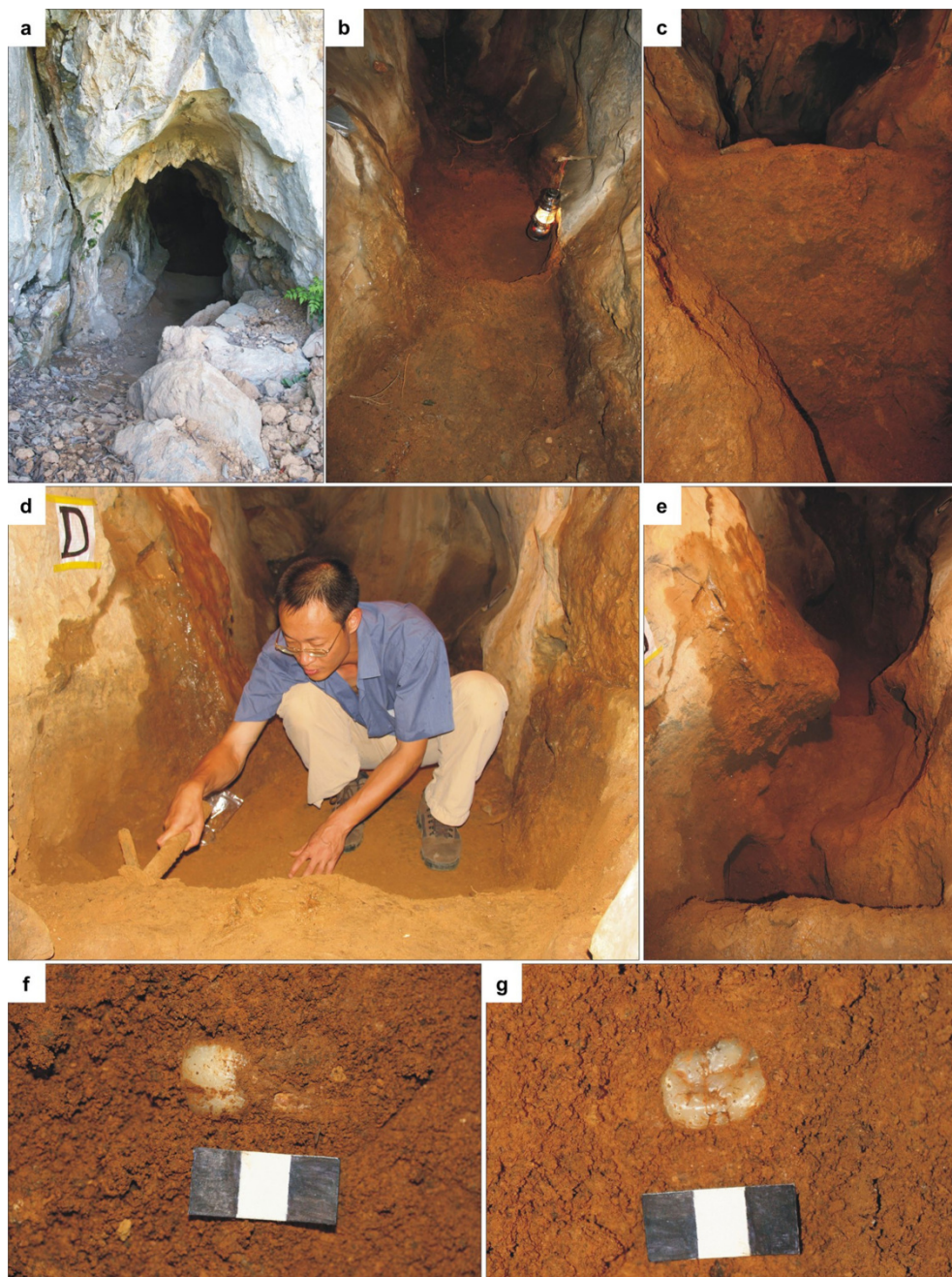
Peer review information *Nature* thanks Benedikt Kessler and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

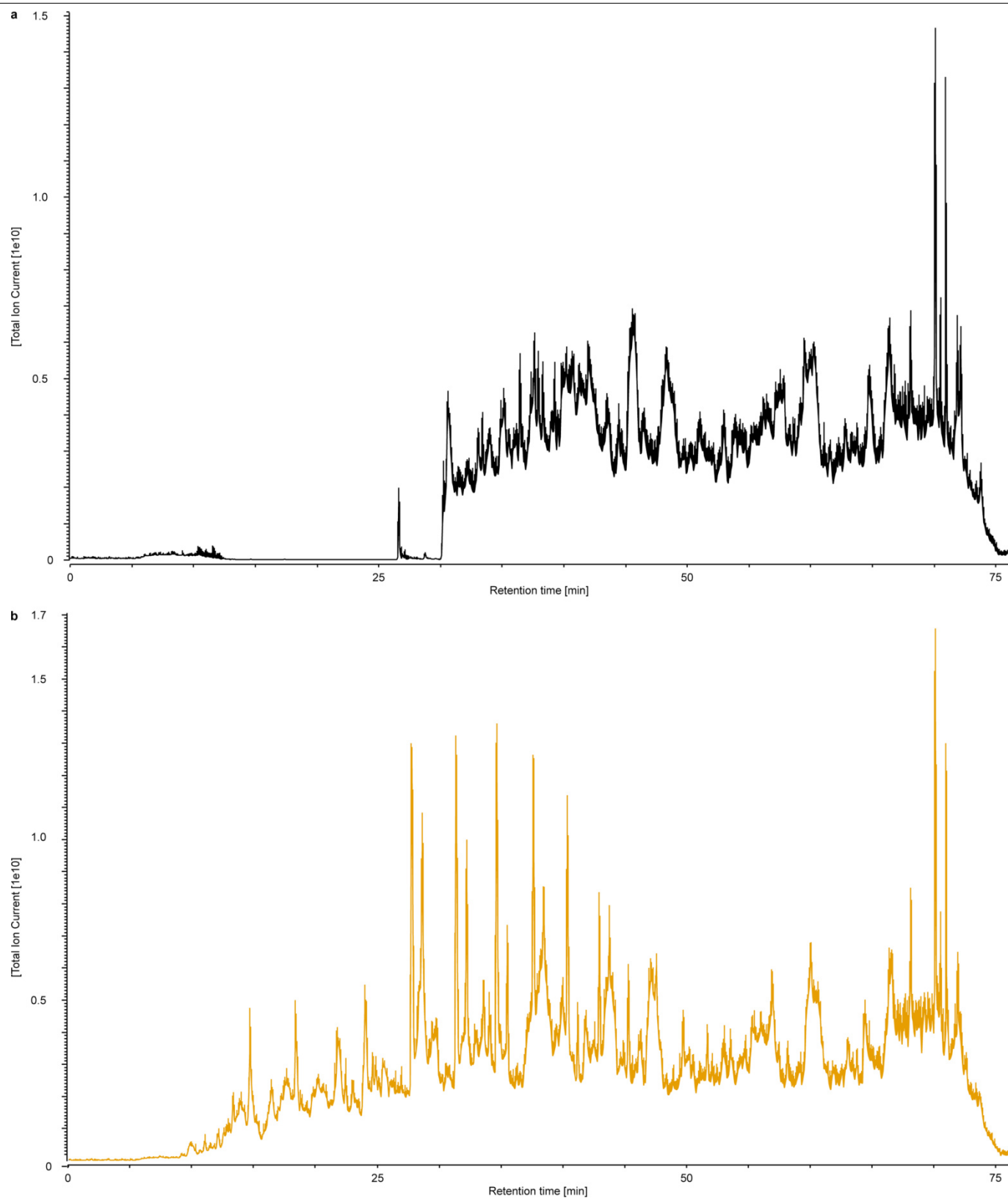


Extended Data Fig. 1|The current environment of Chuifeng Cave.

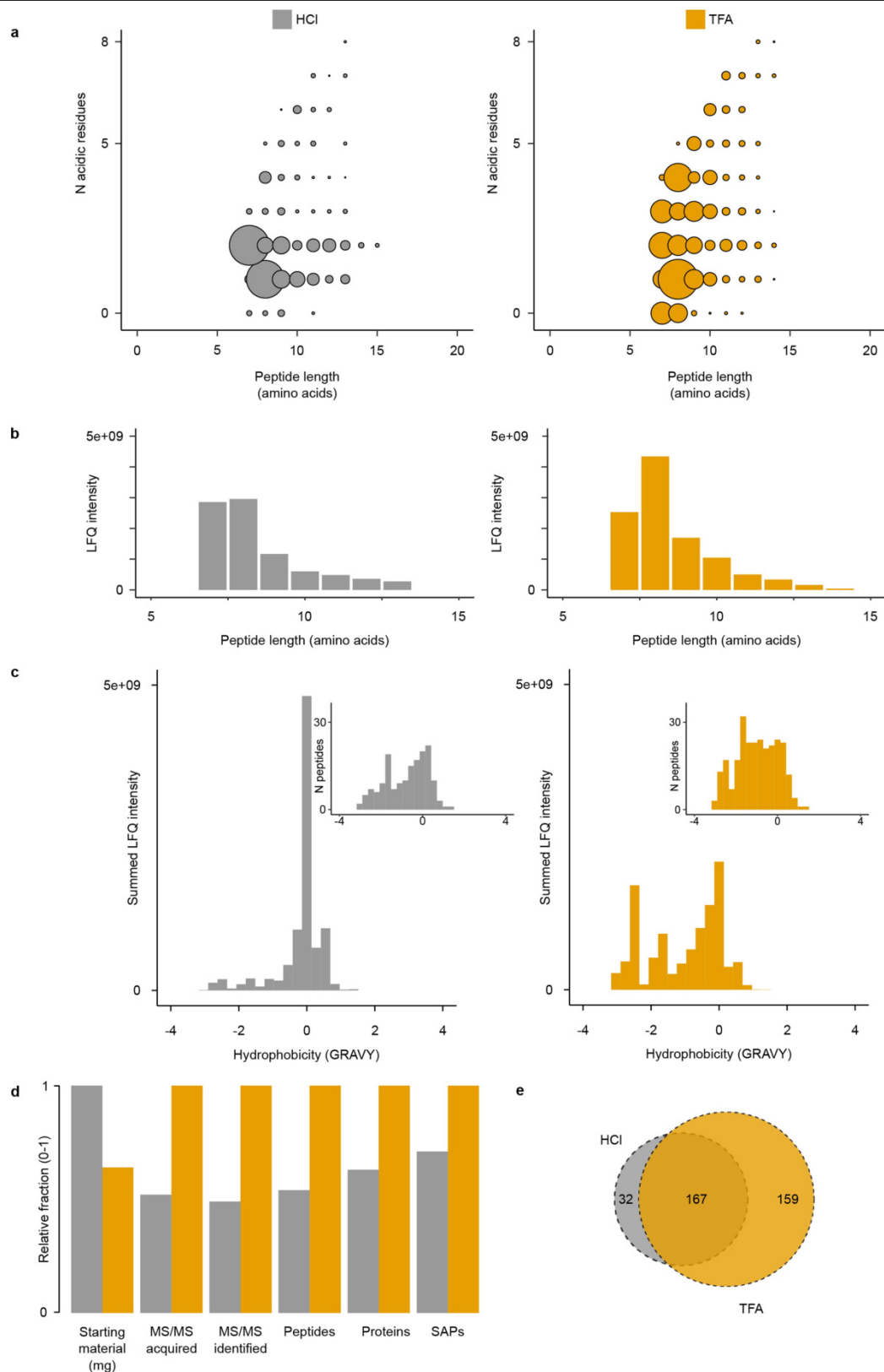
a. Landscape outside Chuifeng Cave. **b.** Elevated altitude of Chuifeng Cave (arrow points to the entrance). Photo credit: W.W.



Extended Data Fig. 2 | Excavations in Chuifeng Cave. **a**, Main entrance of Chuifeng Cave. **b**, Well-preserved deposits before excavation. **c**, The stratigraphic profile (1.3 m in height) of area D. **d**, W.W. excavating in area D. **e**, Excavated channel. **f, g**, In situ *G. blacki* teeth (scale bars, 3 cm). Photo credit: W.W.



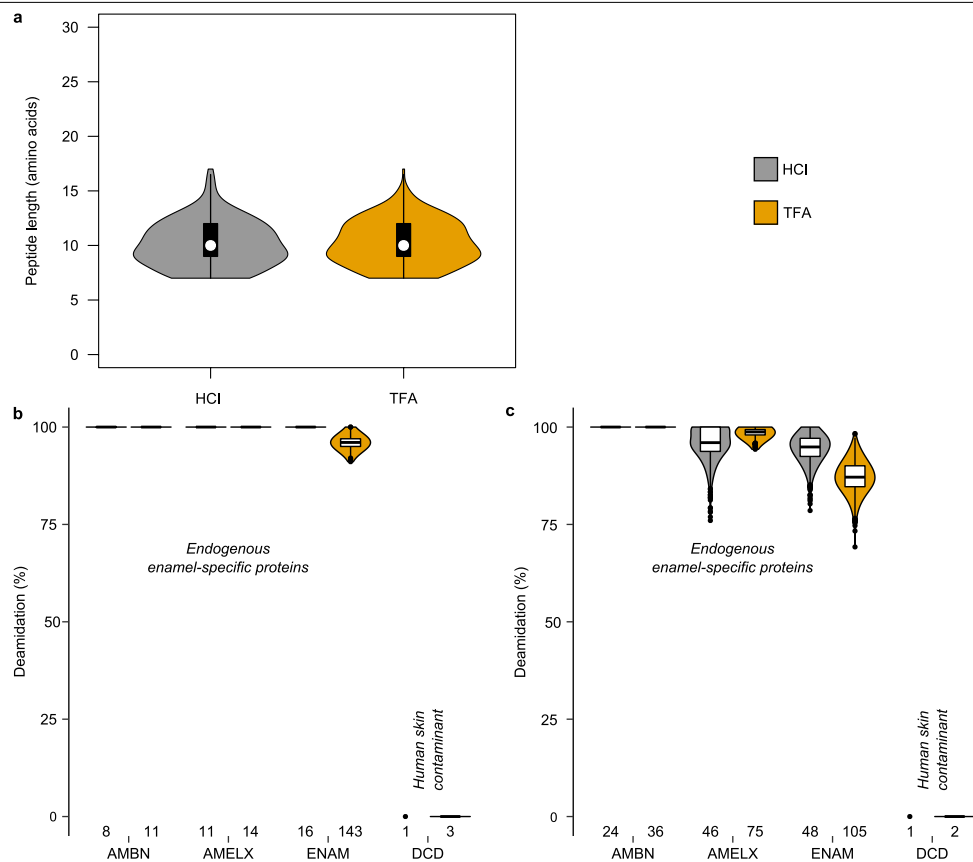
Extended Data Fig. 3 | Total ion current chromatograms of the analysed samples. a, HCl extract. **b,** TFA extract. Note differences in maximum total ion currents on the yaxes. Each extract was analysed only once.



Extended Data Fig. 4 | Comparison of HCl and TFA extraction protocols.

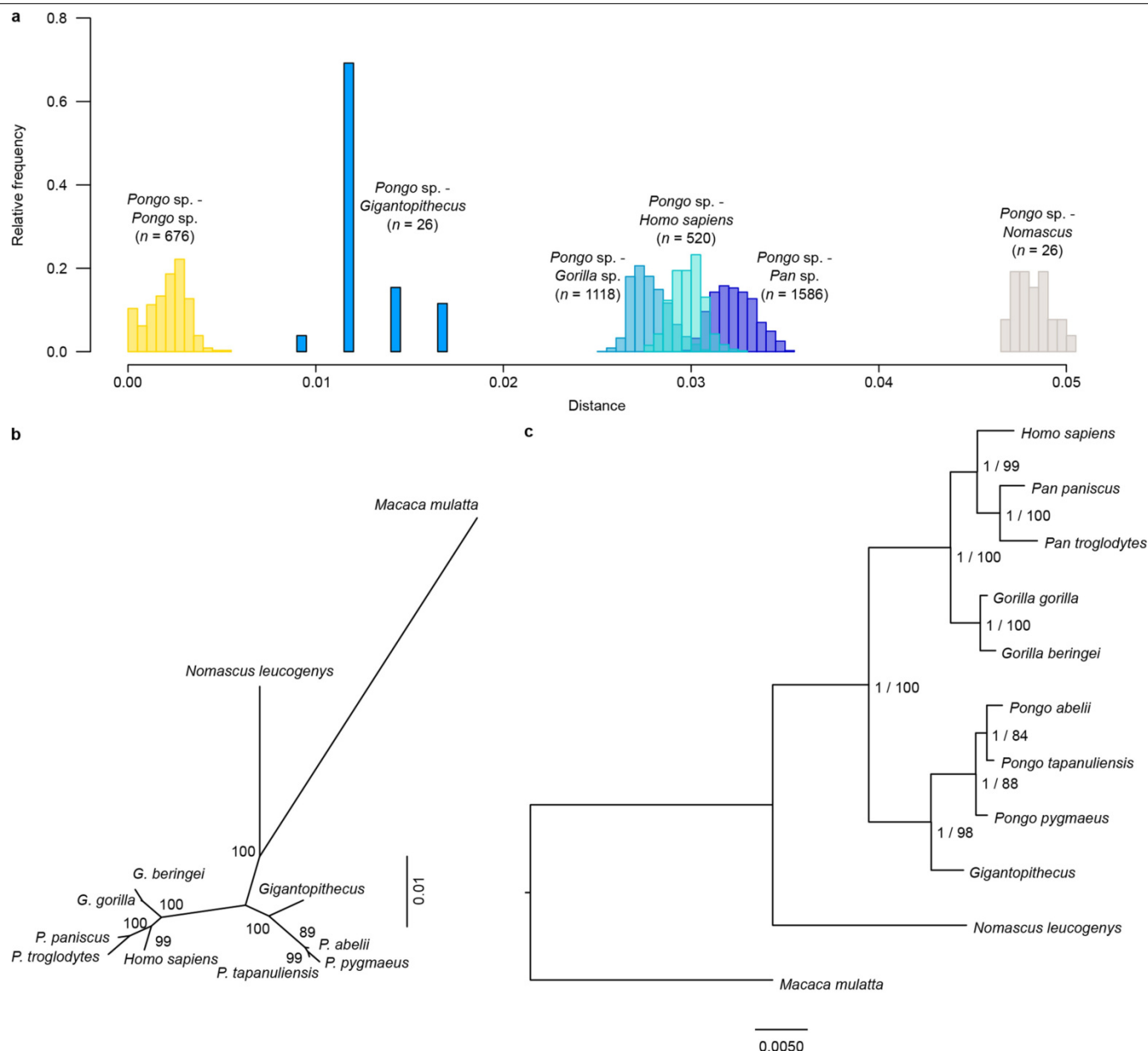
a, Summed and normalized peptide intensities for each combination of peptide length and number of acidic residues (aspartic acid, glutamic acid, deamidated asparagine, deamidated glutamine). Circle sizes are proportional to the percentage of the total intensity, for each combination of peptide length and number of acidic residues. **b**, Summed and normalized intensities by peptide length. **c**, Summed and normalized peptide intensities across peptide hydrophobicity (GRAVY index values calculated using the R package Peptides,

scale 'KyteDoolittle'). Insets show peptide count distribution across peptide hydrophobicity. **d**, Extraction performance for various data categories. Values scaled to one and compared to the best-performing extraction method for each category independently. SAPs refer to those SAPs informative within Hominoidea. **e**, Proportional Venn diagram of unique peptide sequences identified in the two demineralization methods. All comparisons based on MaxQuant LFQ data only. N, number of peptides.



Extended Data Fig. 5 | Damage characteristics of HCl and TFA extraction protocols. **a**, Comparison of peptide lengths, showing an identical distribution for the TFA ($n = 305$) and HCl ($n = 191$) extractions (two-sided t -test, $t_{394} = -0.599$, $P = 0.5495$). **b**, Comparison of asparagine deamidation. **c**, Comparison of glutamine deamidation. **b**, **c**, Violin plots describe the distribution of bootstrap replicates ($n = 1,000$) of intensity-based peptide deamidation³². The number of peptides used for the calculation are shown at the bottom. For some proteins,

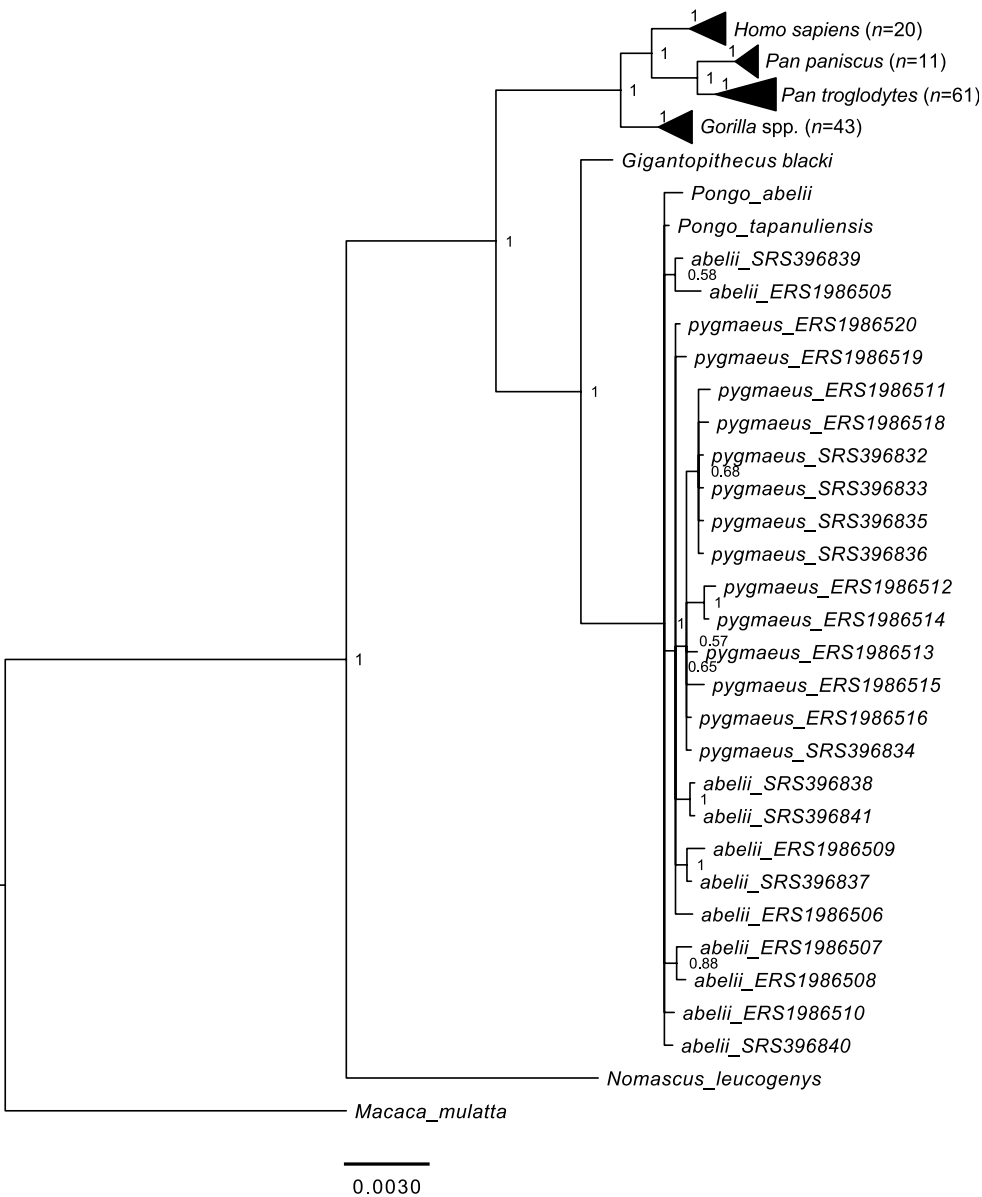
only deamidated asparagines or glutamines were observed (for example, AMBN), while DCD is included as an example of a non-deamidated contaminant. All comparisons based on MaxQuant data only. For **a–c**, box plots define the range of the data (whiskers extending to $1.5 \times$ the interquartile range), outliers (beyond $1.5 \times$ the interquartile range), 25th and 75th percentiles (boxes) and medians (dots).



Extended Data Fig. 6 | Unconstrained phylogenetic analysis of *G. blacki*.

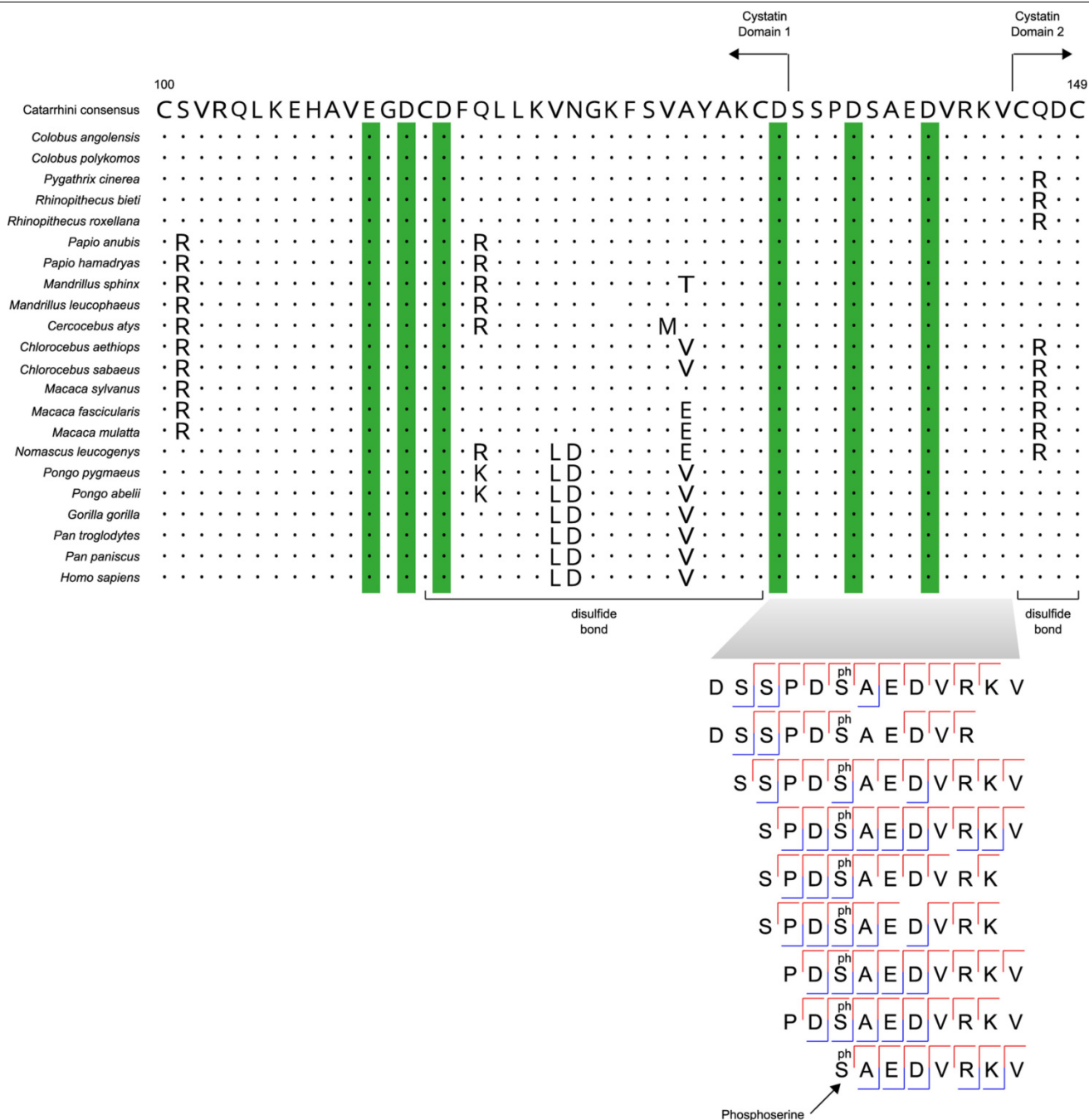
a, Pairwise distances between groups of selected Hominoids and *Pongo* estimated using the concatenated protein alignments and the phangorn R package. *n* shows number of pairwise comparisons. **b**, Maximum-likelihood tree computed on a distance matrix using pml R function. Support values were obtained from 1,000 bootstrap replicates. **c**, Rooted phylogenetic tree

obtained using MrBayes. For each bipartition, we show the posterior probability (0–1) obtained from the Bayesian approach and the support values obtained from 100 non-parametric bootstrap replicates in a PHyML maximum-likelihood (0–100) tree. PHyML and MrBayes recover the same topology. **b** and **c** are based on the same concatenated alignment of the five proteins retrieved from *Gigantopithecus*, and resulted in the same tree topology.



Extended Data Fig. 7 | Bayesian phylogenetic tree of *G. blacki* including 162 modern genomes of great apes and a single hylobatid (*N. leucogenys*). Tree obtained from the concatenated alignment using MrBayes. *M. mulatta* was

used as an outgroup. The internal nodes corresponding to the *Gorilla*, *Pan* and *Homo* clades are collapsed for visualization purposes. The number of individuals in each of these nodes is indicated in parentheses.



Extended Data Fig. 8 | Sequence conservation and structural relevance of retrieved AHSG peptides. All AHSG-specific peptides, identified by PEAKS and MaxQuant, derive from a single sequence region bridging cystatin domains 1 and 2. The surviving sequence region is evolutionarily conserved across Catarrhini. It contains a regular repeat of acidic amino acid residues (aspartic acid, D, on positions 133, 137 and 141) that enable binding of basic calcium phosphate (residues highlighted in green), similarly to a conserved

region just N-terminal (glutamic acid, E, on position 111, and aspartic acid, D, on position 113 and 115). At the bottom, a fragment ion alignment is given of MaxQuant-identified AHSG peptides. The serine is phosphorylated in all matching spectra. The majority-based amino acid consensus sequence for the alignment of Catarrhini is shown at the top for amino acid positions 100–149 (amino acid coordinates following UniProt accession P02765 [FETUA_HUMAN]).

Extended Data Table 1 | Enamel proteome sequence coverage

Protein	Primary entry	Protein accession	MaxQuant peptides (all unique)	MaxQuant amino acids	PEAKS peptides (all unique)	PEAKS amino acids	Combined sequence coverage (%)
AMELX	H2PUX0_PONAB	H2PUX0	149	135 (4)	270	141 (10)	70.7
AMBN	H2PDI5_PONAB	H2PDI5	55	105 (15)	79	107 (11)	27.5
AMTN	H2PDI4_PONAB	H2PDI4	2	18 (0)	2	18 (0)	8.6
ENAM	H2PDI6_PONAB	H2PDI6	125	129 (5)	189	181 (57)	16.3
MMP20	H2NF32_PONAB	H2NF32	2	9 (0)	1	9 (0)	1.9
AHSG	H2PC98_PONAB	H2PC98	7	13 (0)	12	13 (0)	3.5
ALB	ALBU_Bovin	P02769	2				
DCD	DCD_human	P81605	3		8		
B2MG	B2MG_human	P61769	2				
K1C9	K1C9_human	P35527			3		

Only proteins with two unique peptides in at least either MaxQuant or PEAKS searches were accepted. No protein matches in the dentine or blank extractions fulfilled this criterion. Primary entries for proteins used for phylogenetic reconstruction refer to the *P. abelii* accessions in UniProt for reference purposes. Protein sequence coverage in the final column indicates the coverage obtained after combining PEAKS and MaxQuant peptide recovery and is only reported for proteins considered endogenous. ALB, DCD and K1C9 are considered to be contaminants. For amino acid columns, numbers in brackets refer to amino acid positions uniquely identified in PEAKS or MaxQuant searches.

Extended Data Table 2 | MAT estimation at Chuifeng Cave

WMO station ID	WMO station name	Longitude	Latitude	WMO altitude (m)	Estimated altitude (m)	First month	Last month	MAT	Altitude source used	Chuifeng Cave MAT
5920901	Ching His	106.42E	23.13N	740	743	February 1981	October 1990	19.5	WMO	22.8
5698505	Hekou	103.95E	22.5N	137	114	January 1961	December 1970	22.6	WMO	22.1
5791601	Tien-O	107.17E	25.0N	305	245	January 1981	October 1990	19.7	WMO	20.0
5904602	Ta Wan	109.42E	23.85N	76	81	January 1981	October 1990	20.8	WMO	20.0
5963201	Tung Hsing	107.97E	21.55N	13	10	February 1981	October 1990	23.0	WMO	22.0
5921100	Bose	106.6E	23.9N	na	154	January 1961	October 1990	22.5	Estimated altitude	22.2
5920901	Napo	105.95E	23.3N	na	1214	January 1981	October 1990	19.6	Estimated altitude	24.5
5900700	Guangnan	105.07E	24.07N	na	1257	January 1981	October 1990	17.6	Estimated altitude	22.8
5943100	Nanning	108.35E	22.82N	na	81	January 1922	November 1993	22.0	Estimated altitude	21.2
5902300	Hechi	108.05E	24.7N	na	204	January 1981	October 1990	21.2	Estimated altitude	21.1

The ten geographically closest meteorological weather stations included in publicly available WMO data were used to estimate current MAT at Chuifeng Cave. Correlation between online altitude estimation and WMO provided altitude is $R^2 = 0.99$ (Pearson correlation).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Mass spectrometric data were acquired using the Xcalibur™ Software, controlling the Thermo Scientific™ LC-MS systems.

Data analysis

Xcalibur™ (version 4.1)
MaxQuant (version 1.6.20)
PEAKS (version 7.5)
Geneious (version 5.4.4)
Phangorn (version 2.4.0, R version 3.4.2)
BWA-MEM (version 0.7.7)
PICARD (version 1.91)
GATK UnifiedGenotyper (version 3.4-46)
blastall (version 2.2.26)
PhyloBot (version 10.09.2016.1)
PHyML (version 3.1)
MrBayes (version 3.2.6)
mafft (version 7.205)
R (version 3.4.3)
R, package Peptides (version 2.4)
R, package vioplot (version 0.2)

R, package ggplot2 (version 3.1.0)
R, package eulerr (version 5.1.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the data set identifier PXD013838. Protein sequences generated as part of this study are provided as supplementary files.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was required. All available Gigantopithecus specimen samples (n=1) were analyzed.
Data exclusions	No data was excluded from the study.
Replication	Phylogenetic trees were reproduced using three different algorithms, and found consistent results (see Methods and SI). Proteomic results were replicated using repeated LC-MS/MS runs of two different extracts for enamel, and we observed consistent results within and between those two extracts.
Randomization	Samples were injected in the LC-MS/MS system in randomised order.
Blinding	Blinding was not relevant to this study, as only a single specimen was analyzed

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Palaeontology

Specimen provenance	The studied specimen derives from the Chufeng Cave palaeontological site in China (see Methods). Export of specimens to the Department of Biology, University of Copenhagen was regulated by permit #2018-01 issued by Prof. Wang Wei. Wei, co-author and PI of the excavation and research of Chufeng Cave, director of the Anthropology Museum of Guangxi, China (issued on February 26, 2018, by the Anthropology Museum of Guangxi, China).
---------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Specimen deposition

Specimens are available upon request to E. Cappellini (Department of Biology, University of Copenhagen, Denmark), or W. Wang (Shandong University, China).

Dating methods

No new dates obtained.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Distributed coding of choice, action and engagement across the mouse brain

<https://doi.org/10.1038/s41586-019-1787-x>

Nicholas A. Steinmetz^{1,3*}, Peter Zátka-Haas², Matteo Carandini^{1,4} & Kenneth D. Harris^{2,4}

Received: 21 November 2018

Accepted: 10 October 2019

Published online: 27 November 2019

Vision, choice, action and behavioural engagement arise from neuronal activity that may be distributed across brain regions. Here we delineate the spatial distribution of neurons underlying these processes. We used Neuropixels probes^{1,2} to record from approximately 30,000 neurons in 42 brain regions of mice performing a visual discrimination task³. Neurons in nearly all regions responded non-specifically when the mouse initiated an action. By contrast, neurons encoding visual stimuli and upcoming choices occupied restricted regions in the neocortex, basal ganglia and midbrain. Choice signals were rare and emerged with indistinguishable timing across regions. Midbrain neurons were activated before contralateral choices and were suppressed before ipsilateral choices, whereas forebrain neurons could prefer either side. Brain-wide pre-stimulus activity predicted engagement in individual trials and in the overall task, with enhanced subcortical but suppressed neocortical activity during engagement. These results reveal organizing principles for the distribution of neurons encoding behaviourally relevant variables across the mouse brain.

Performing a perceptual decision involves processing sensory information, selecting actions that may lead to reward, and executing these actions. It remains unknown how the neurons mediating these processes are distributed across brain regions, and whether they rely on circuits that are shared or distinct. Most studies of action selection (hereafter referred to simply as choice) have focused on individual regions, such as frontal, parietal and motor cortex, basal ganglia, thalamus, cerebellum and superior colliculus^{4–11}. However, neural correlates of movements, rewards, and other task variables have been observed in multiple brain regions, including in areas previously identified as purely sensory^{12–24}. It is therefore possible that many brain regions also participate in action selection. Nevertheless, neuronal signals that correlate with action do not necessarily correlate with choice. To carry choice-related signals, a brain region must contain neurons whose firing selectively predicts the chosen action before the action occurs²⁵.

Successful performance in a perceptual task depends not only on choosing the correct action, but also on choosing to engage in the task in the first place. Stimuli that drive actions during an engaged behavioural state do not necessarily drive actions when disengaged; for example, in contexts where the action will not lead to reward. Furthermore, even well-trained participants often show varying levels of behavioural engagement or vigilance within a task, resulting in varying probability of responding promptly and accurately to sensory stimuli^{26–28}. At times of low engagement, stimuli arriving at the sense organs evidently fail to effectively drive the circuits responsible for selecting and initiating action. It remains unclear whether this context-dependent gating occurs globally²⁹ or whether it involves multiple brain systems differentially³⁰.

Brain-wide recording in visual behaviour

To determine the distribution of neurons encoding vision, choice, action, and behavioural engagement, we recorded neural activity across

the brain while mice performed a task that enabled us to distinguish these processes (Fig. 1a–c). This task combines the advantages of two-alternative forced choice and Go–NoGo designs^{3,31}. On each trial, visual stimuli of varying contrast could appear on the left side, right side, both sides or neither side. Mice earned a water reward by turning a wheel with their forepaws to indicate which side had highest contrast (Fig. 1a–c). If neither stimulus was present, they earned a reward for making a third type of response: keeping the wheel still for 1.5 s. If left and right stimuli had equal non-zero contrast, the mice were rewarded randomly for left or right turns. The same visual stimulus could therefore lead to either direction of turn, or to no action, enabling us to dissociate the neural correlates of visual processing, of action initiation (turning the wheel versus holding it still) and of action selection (turning left versus right).

Mice performed the task proficiently (Fig. 1c, Extended Data Fig. 1). Their choices were most accurate when stimuli appeared on a single side at high contrast ($1.7 \pm 2.5\%$ incorrect choices, defined as turns in the wrong direction; $10.1 \pm 8.3\%$ Misses, defined as failures to turn; mean \pm s.d., $n = 39$ sessions, 10 mice). They performed less accurately in more challenging conditions: with low-contrast single stimuli ($5.1 \pm 6.4\%$ incorrect choices; $29.8 \pm 19.8\%$ Misses); or with competing stimuli of similar but unequal contrast ($20.0 \pm 7.9\%$ incorrect choices; $13.9 \pm 11.7\%$ Misses, on trials with high versus medium or medium versus low contrast). As expected in these more challenging cases, reaction times were longer (Fig. 1d; $P < 10^{-4}$, multi-way ANOVA).

While mice performed the task, we used Neuropixels probes^{1,2} to record from approximately 30,000 neurons in 42 brain regions (Fig. 1e–j; abbreviations of brain regions are defined in Extended Data Table 1). Inserting two or three probes at a time in the left hemisphere yielded simultaneous recordings from hundreds of neurons in multiple regions during each recording session ($n = 92$ probe insertions over 39 sessions in 10 mice, Fig. 1h, i). We identified the firing times of individual neurons using Kilosort³² and phy³³, and determined their anatomical locations by combining

¹Institute of Ophthalmology, University College London, London, UK. ²Institute of Neurology, University College London, London, UK. ³Present address: Department of Biological Structure, University of Washington, Seattle, WA, USA. ⁴These authors jointly supervised this work: Matteo Carandini, Kenneth D. Harris. *e-mail: nick.steinmetz@gmail.com

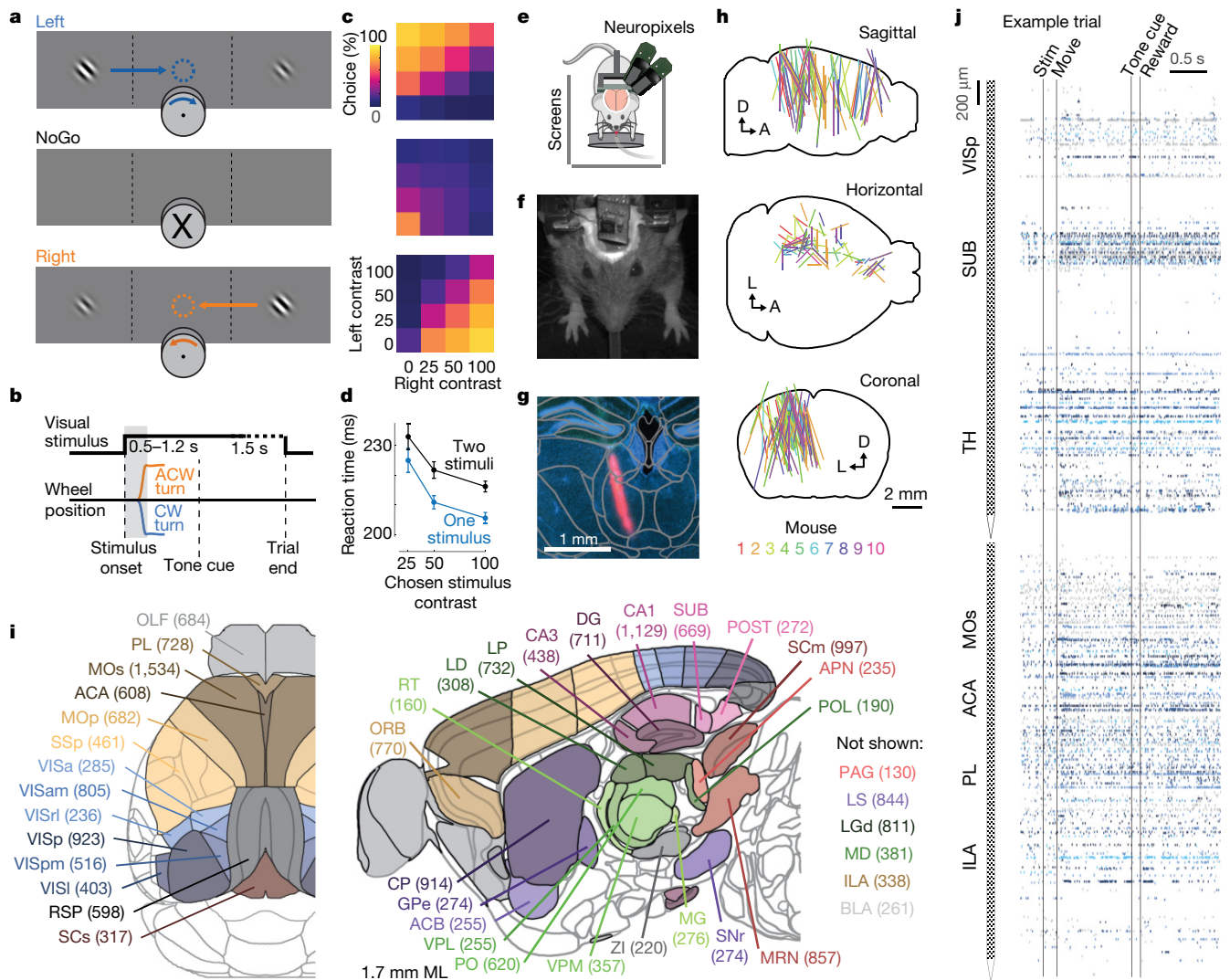


Fig. 1 | Brain-wide recordings during a task that distinguishes vision, choice and action. **a**, Mice earned water rewards by turning a wheel to indicate which of two visual gratings had higher contrast, or by not turning if no stimulus was presented. When stimuli had equal contrast, a left or right choice was rewarded with 50% probability. Grey rectangles indicate the three computer screens surrounding the mouse. Arrows (not visible to the mouse) indicate the rewarded wheel turn direction and the coupled movement of the visual stimulus (X indicates reward for no turn), and the coloured dashed circle (not visible to the mouse) indicates the stimulus location at which a reward was delivered. **b**, Timeline of the task. Mice were free to move as soon as the stimulus appeared, but the stimulus was fixed in place and rewards were unavailable until after an auditory tone cue. If no movement was made for 1.5 s after the tone cue, a NoGo was registered. The grey region is the analysis window, from 0 to 0.4 s after stimulus onset. CW, clockwise; ACW, anticlockwise. **c**, Average task performance across subjects; $n = 10$ mice, 39 sessions, 9,538 trials. Colour maps depict the probability of each choice given the combination of contrasts presented. **d**, Reaction time as a function of stimulus contrast and presence of

competing stimuli. **e**, Mice were head-fixed with forepaws on the wheel while multiple Neuropixels probes were inserted for each recording. **f**, Frontal view of subject performing the behavioural task during recording, with forepaws on wheel and lick spout for acquiring rewards. **g**, Example electrode track histology with atlas alignment overlaid. **h**, Recording track locations as registered to the Allen Common Coordinate Framework 3D space. Each coloured line represents the span recorded by a single probe on a single session, coloured by mouse identity. D, dorsal; A, anterior; L, left. **i**, Summary of recording locations. Recordings were made from each of the 42 brain regions coloured on the top-down view of cortex (left) and sagittal section (right). For each region, the number in parentheses indicates total recorded neurons. For abbreviations, see Extended Data Table 1. **j**, Spike raster from an example individual trial, in which populations of neurons were simultaneously recorded across visual and frontal cortical areas, hippocampus and thalamus. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (version 3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).

electrophysiological features with histological reconstruction of fluorescently labelled probe tracks (Fig. 1g, Extended Data Figs. 2, 3). Across all sessions, we recorded from 29,134 neurons ($n = 747 \pm 38$ neurons per session, mean \pm s.e.), of which 22,458 were localizable to one of 42 brain regions.

Propagation of activity during the task

Trial onset was followed by increased average activity in nearly all recorded regions. A sizeable fraction of all neurons (60.0%, 13,466

neurons) showed significant modulation of firing rate during the task ($P < 0.05$; Bonferroni-corrected test for six task epochs; Supplementary Figs. 1, 2; Methods). Most of these neurons (74.3%) consistently increased their activity during the task, but a sizeable minority (20.2%) consistently decreased their activity. Neurons were diverse in the timing of their activity during trials, with timing differences both within and between brain areas (Extended Data Fig. 4a, b); however, neuronal activity was detectable before the onset of wheel movement in most regions (Extended Data Fig. 4c, d). Similarly widespread

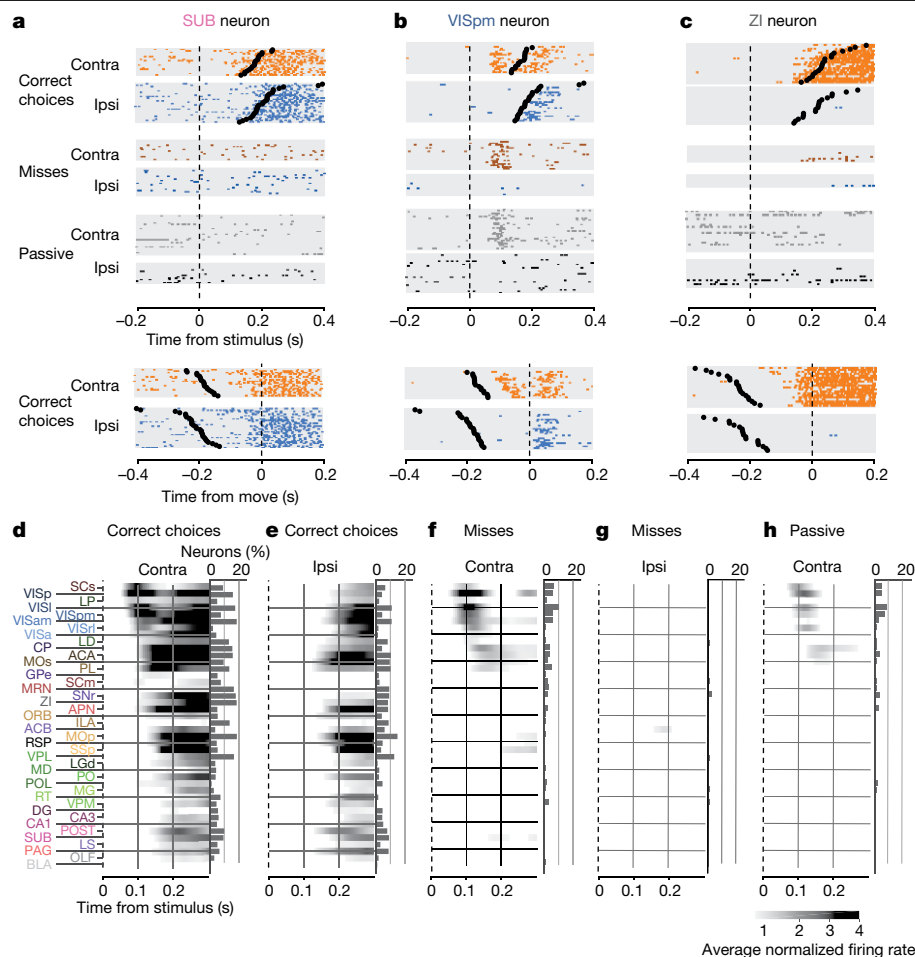


Fig. 2 | Activity propagates from a visual pathway to the entire brain during task performance. **a–c**, Rasters showing activity of three example neurons following visual stimuli presented on the contralateral or ipsilateral side alone, on correct choice trials (when they evoked wheel turns in the correct direction), Miss trials (when mice failed to respond in the task context) and when stimuli were presented in a passive context with no opportunity to earn reward. Top six panels: aligned to stimulus onset, black dots represent movement. Bottom two

panels: aligned to movement, black dots represent stimulus onset. **d–h**, Colour maps showing firing rates averaged over responsive neurons in each region and over trials of the indicated type. Contralateral visual stimulus contrasts were matched between **d**, **f** and **h** so that differences in activity do not reflect differing visual drive. Subpanels to the right of each colour map represent the percentage of neurons in each area significantly more responsive during that condition than baseline ($P < 10^{-4}$; see Methods).

activity was observed following reward delivery (Extended Data Fig. 4e).

Examining rasters of individual neurons' activity across trials revealed consistent correlates of action initiation, sensory stimuli or choices (Fig. 2a–c). For example, a neuron in the subiculum (Fig. 2a) produced no response to the visual stimuli, but consistently fired before wheel turns regardless of their direction. Such non-specific movement correlates were also often seen in neurons that produced sensory responses. For example, a neuron in visual cortex (Fig. 2b) showed activity following onset of visual stimulus that was selective for stimulus location but also fired following wheel turns, regardless of the subject's choice (that is, direction of wheel turn). Neurons with choice-selective responses were rare but could be found in select nuclei: for example, a neuron in the zona incerta (ZI; Fig. 2c) showed no visual response but increased its firing rate before contralateral choices, with no response before or after ipsilateral choices.

Throughout the brain, most of the activity following trial onset reflected non-specific movement correlates (Fig. 2d–h). When a mouse successfully selected a visual stimulus contralateral to the recorded hemisphere, activity emerged first in classical visual regions such as visual cortex (VIS) and superficial superior colliculus (SCs), and soon spread to most of the remaining recorded regions (Fig. 2d). When the mouse successfully selected an ipsilateral stimulus, most areas

were again activated, but VIS and SCs were now among the last areas to respond, rather than the first (Fig. 2e). When the mouse missed a contralateral stimulus, leading to no action, activity was found in a 'visual pathway' consisting of classical visual areas, basal ganglia and several midbrain structures (Fig. 2f), but failed to propagate globally. When the mouse missed an ipsilateral stimulus, however, the recorded hemisphere remained largely silent (Fig. 2g). The widely distributed activity seen following trial onset was therefore present only when mice moved, regardless of the particular stimulus and particular action.

Outside of the task context, responses to visual stimuli were similar to those of Miss trials, but generally weaker (Fig. 2h). We measured activity in passive replay periods following task performance, when the same stimuli were presented without the opportunity to earn rewards. In these passive trials, the mice hardly ever turned the wheel ($94.1\% \pm 0.6\%$ of trials with high contrast stimuli had no movement). Stimuli contralateral to the recorded hemisphere gave rise to weak activity restricted to the visual pathway (Fig. 2h). No activity was seen on average following passive presentation of stimuli ipsilateral to the recorded hemisphere (not shown).

In sum, this analysis of average activity suggests that whereas responses to visual stimuli are largely confined to a restricted visual pathway, neural correlates of action initiation are essentially global. To assess the distribution of these signals at a finer scale and to search for

signals encoding choice, we next examined the activity of individual neurons.

Globally distributed action coding

To analyse the firing correlates of individual neurons, we used an approach based on kernel fitting (Extended Data Fig. 5). We fit the activity of each neuron with a sum of kernel functions time-locked to stimulus presentation and to movement onset. We fit six stimulus-locked kernels—one for each of three possible contrast values on each side ('Vision' kernels), which captured variations in amplitude and timing of the visual activity driven by different stimuli. We fit two movement-locked kernels: an 'Action' kernel triggered by a movement in either direction and a 'Choice' kernel capturing differences in activity between left and right movements.

To determine which neurons encoded vision, action and choice, we used a nested test: we fit a model including all kernels except the one to be tested and determined whether adding the test kernel improved the model's prediction of held-out data. Applying this test to the example neurons from before, we find that this method succeeded in quantifying the contralateral visual stimulus (Fig. 3a) and action (Fig. 3b) correlates inferred from examining each trial type (compare with Fig. 2a, b). In determining whether a neuron passed this test, we used parameters that gave false-positive error rates of 0.33% on shuffled data (Extended Data Fig. 5h). As our question concerns activity predictive of upcoming movements, we applied this analysis only to pre-movement activity. Consistent with its raster plot (Fig. 2a), the example subicular neuron examined earlier required only an Action kernel, indicating entirely non-selective action correlates (Fig. 3b). By contrast, the example visual cortical neuron (Fig. 2b) required only Vision kernels (Fig. 3a), indicating that it had exclusively visual correlates before action initiation. The fraction of cross-validated variance explained by the kernels was frequently small (Fig. 3c, Extended Data Fig. 6a), even for neurons whose mean rates they accurately predicted (50.2% for the neuron in Fig. 3a and 13.6% for the neuron in Fig. 3b), as expected from trial-to-trial variability and encoding of task-independent variables^{14,34}.

Neurons encoding vision (that is, requiring Vision kernels) were found in a pathway comprising primarily classical visual areas (Fig. 3d, f). They were common in VIS, thalamus and superficial superior colliculus (SCs), but were also occasionally present in other structures such as frontal cortex (MOs, ACA and PL), basal ganglia (CP, GPe and SNr), and several midbrain nuclei (SCm, MRN, APN and ZI; Fig. 3d, f, Extended Data Fig. 6a; brain regions are listed in Extended Data Table 1).

By contrast, neurons encoding action (requiring an Action kernel) were spread throughout all recorded regions (Fig. 3e, g). The distribution of these neurons encoding action was significantly broader than that of neurons encoding visual stimuli (Extended Data Fig. 4f). A large majority of neurons encoding action did not require an additional Choice kernel—they responded equally for movements in either direction. The rare exceptions requiring a Choice kernel are discussed next.

Choice coding in forebrain and midbrain

Neurons encoding specific choices were found in a small subset of brain regions (Fig. 4a, b). We identified choice-selective neurons as neurons for which the Choice kernel was required to explain their activity in the nested test described above. These neurons were rare and were distributed across frontal cortex (MOs, PL and MOp), basal ganglia (CP and SNr), higher-order thalamus and motor-related superior colliculus (SCm), as well as in two unexpected subcortical nuclei, the midbrain reticular nucleus and ZI (Fig. 4b). This set of regions encoding choice overlapped partially with the visual pathway—both included frontal cortex, basal ganglia and several midbrain structures, but choice-selective neurons were not found in VISp. Neurons encoding choice were again significantly more localized than neurons encoding action (Extended

Data Fig. 4f). To further confirm these conclusions, we developed a version of choice-probability analysis for tasks with many stimulus conditions, called 'combined-conditions choice probability' (Methods). This statistic quantifies the probability that a neuron's spike count will be greater on trials with one choice than with other choices, for matched stimulus conditions as in classical choice probability. This analysis produced similar results (Extended Data Fig. 7).

Across choice-encoding regions, choice signals emerged with similar timing (Fig. 4c, d). We trained a decoder to predict the subject's choice from recorded population activity after first subtracting the prediction of population activity from the Vision and Action kernels (to yield a decoding of choice isolated from visual and non-specific action signals). This population-level decoding identified similar areas encoding each variable as did the individual neuron decoding (Extended Data Fig. 5g), and we found that the time course of choice decoding was not significantly different across choice-selective regions in frontal cortex, striatum and midbrain (two-way ANOVA on brain region and time, interaction $P > 0.05$; Fig. 4c). We validated this conclusion using joint peri-event canonical correlation (jPECC) analysis, an extension of the 'joint peristimulus time histogram' method^{35,36}, modified to detect correlations in a 'communication subspace'³⁷ between two populations. Whereas jPECC analysis revealed a consistent time lag for activity correlations between visual and frontal cortex (and between visual cortex and midbrain choice areas), it revealed no lag for activity correlations between frontal and midbrain areas (Fig. 4d, Extended Data Fig. 8).

Although the encoding of choice emerged with indistinguishable timing in midbrain and forebrain, these regions encoded choice differently (Fig. 4e–h). In midbrain (MRN, SCm, SNr and ZI), nearly all choice-selective neurons (53 out of 54, 98%) preferred contralateral choices (Fig. 4e, top and Fig. 4f). By contrast, choice-selective neurons in forebrain (MOs, PL, MOp and CP) could prefer either choice, with a sizeable proportion preferring ipsilateral choices (19 out of 48, 40%), significantly more than in the midbrain ($P < 10^{-5}$, Fisher's exact test; Fig. 4e, bottom and Fig. 4g). Moreover, many midbrain choice-selective neurons exhibited directionally opposed activity: their activity increased before one choice and decreased below baseline before the other (29 out of 54, 54%; note points to the left of $x = 0$ in Fig. 4f). By contrast, neurons in the forebrain typically increased firing before both left and right choices (10 out of 48, 21% suppressed for non-preferred choice, significantly less than in the midbrain; $P < 10^{-3}$, Fisher's exact test; Fig. 4g). Neurons encoding choice, therefore, exhibit a distinctive bilateral encoding of both choices in the forebrain, versus a unilateral encoding of contralateral choices in the midbrain (Fig. 4h).

Distributed coding of task engagement

We next investigated whether engagement in a trial or in the overall task corresponded to characteristic patterns of brain activity. We reasoned that Go trials (when the mouse made a left or right choice), Miss trials and passive visual responses (measured outside the task) might represent three points along a continuum corresponding to progressively lower levels of task engagement.

We began by comparing two conditions in which visual stimuli and behavioural reports were identical: passive visual responses measured outside the task, and responses on Miss trials during the task (Fig. 5a–c). Even though the two conditions were matched for visual stimulation and (lack of) action, they were accompanied in many areas by different activity, both before and after stimulus presentation (Fig. 5a, Extended Data Fig. 6b). Consistent with the average firing rates presented earlier (Fig. 2f, h), more neurons were significantly activated by visual stimuli during the task (Miss trials), than during passive stimulation (Fig. 5b). Differences were also seen in pre-stimulus firing rates; for instance, pre-stimulus activity in VISp was lower in the task (Miss trials) than during passive stimulation, whereas pre-stimulus activity in CP showed the opposite modulation (Fig. 5a). Whereas neocortex and sensory

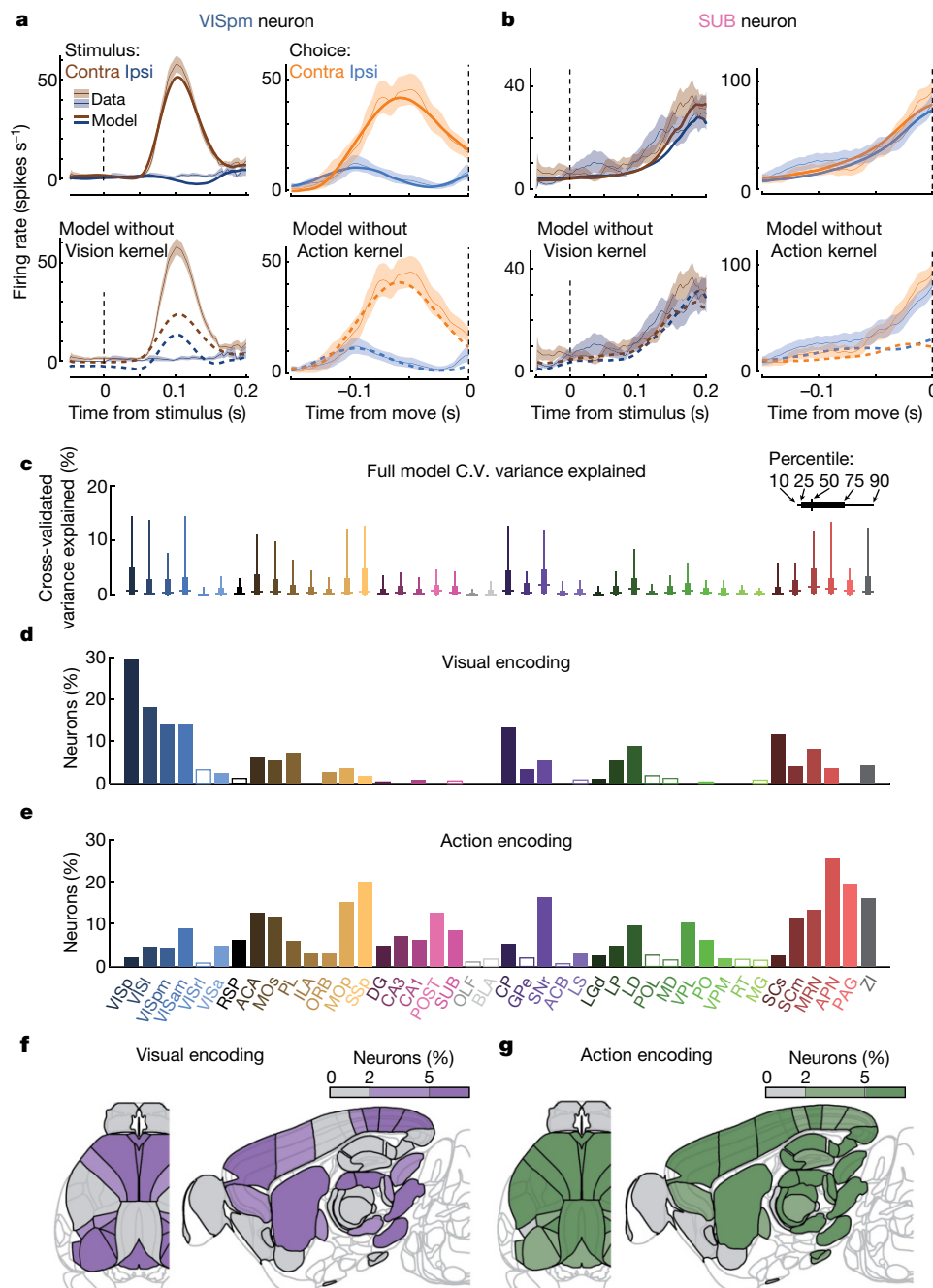


Fig. 3 | Neurons encoding vision are localized but neurons encoding action are found globally. **a**, Example of regression analysis for the example VISpm neuron shown in Fig. 2b. Firing rate was averaged (solid thin line, mean; shaded regions, s.e.m. across trials) across the trial types indicated: all trials with contralateral or ipsilateral stimuli (brown or blue; left panels), and all trials with contralateral or ipsilateral choices (orange or blue, right panels). Top graphs show mean firing rate (thin lines and shading) overlaid with cross-validated prediction of the regression model using all kernels (solid thick lines). Bottom graphs show mean rate (thin lines and shading) overlaid with fits excluding the indicated kernel (dashed lines). The good fit of the full model is lost when excluding the contralateral Vision kernels, indicating that this neuron has stimulus-locked activity that cannot be explained by other variables. **b**, Similar

analysis for activity of the SUB neuron from Fig. 2a, for which a good fit cannot be obtained without the Action kernel. **c**, Box plots showing distribution of the percentage of spiking variance explained in cross-validated (C.V.) tests of the full model, for all neurons within each brain region. **d**, Fraction of neurons in each brain region for which accurate prediction of pre-movement activity required the contralateral Vision kernel. Empty bars indicate those for which the number of neurons passing analysis criteria was less than 5. **e**, as in **d** but for the Action kernel. **f**, Illustration of **d** on a brain map. White areas were not recorded. **g**, As in **f** but for the Action kernel. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (version 3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).

thalamus showed a net decrease in pre-stimulus activity during task context, other regions—including basal ganglia and other subcortical choice-encoding areas—showed a consistent increase (Fig. 5c, Extended Data Fig. 9a). This effect extended to neurons that were not otherwise responsive during the task (Supplementary Fig. 3).

Consistent with the hypothesis of a continuum of engagement across passive, Miss and Go trials, the pattern of firing accompanying task engagement predicted successful performance on individual trials (Fig. 5d–f). For this analysis, we examined only pre-stimulus activity, which could not be conflated with the large non-specific responses

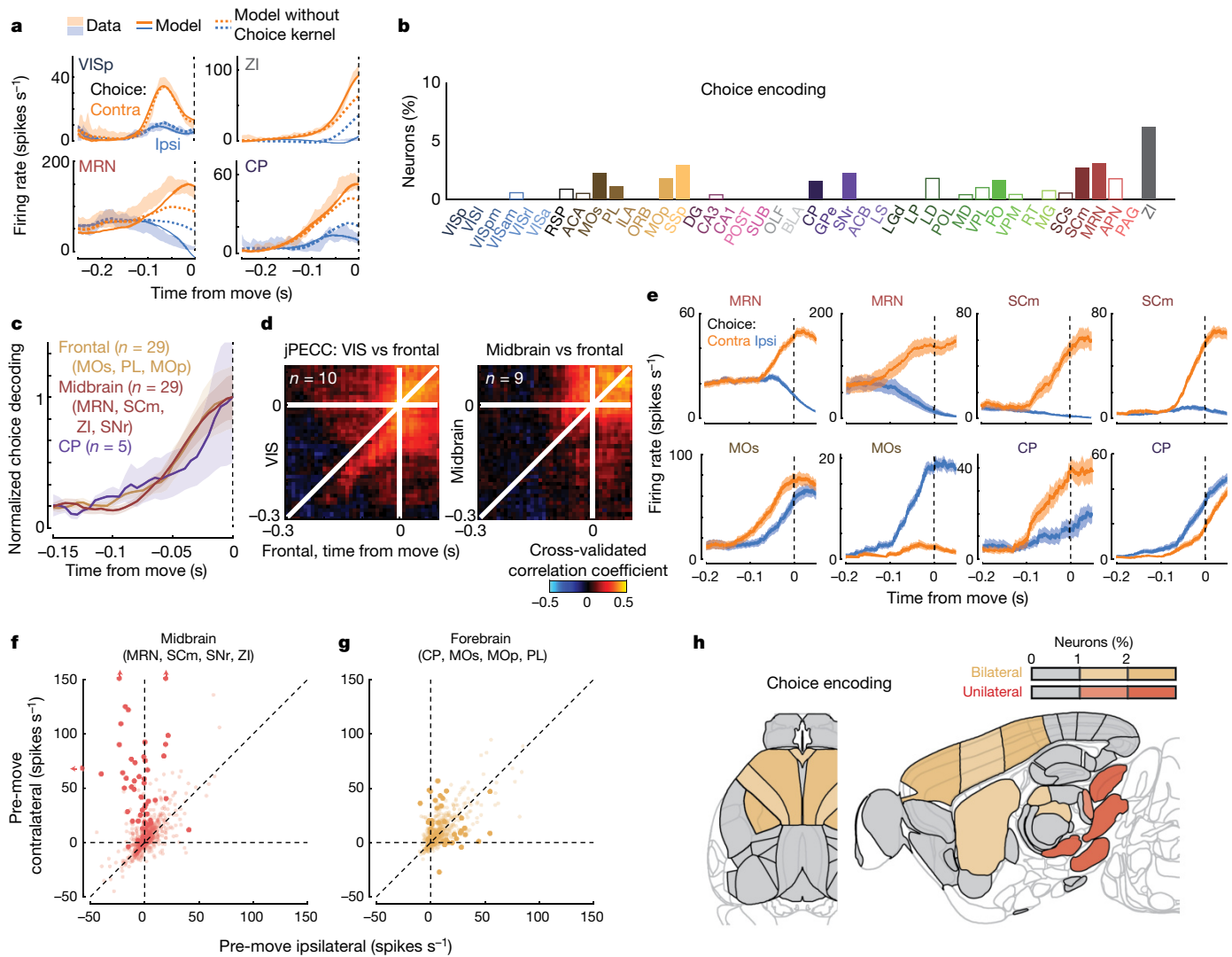


Fig. 4 | Choice signals emerge simultaneously across a localized set of forebrain and midbrain areas. **a**, Firing rates of four example neurons, averaged across indicated trial types (shaded regions, s.e.m. across trials), cross-validated fits of the kernel model using all kernels (solid lines), and fits using all kernels except Choice (dashed lines). The VISp neuron can be accurately predicted without the Choice kernel, indicating that the differing responses between left and right choices can be explained by visual responses. The other three neurons cannot be predicted without the Choice kernel. The ZI neuron also appeared in Fig. 2c. **b**, Fraction of neurons in each brain region for which accurate prediction required the Choice kernel (false-positive rate on shuffled data, 0.3%). Empty bars indicate areas for which the number of neurons passing analysis criteria was less than 5. **c**, Time courses of population decoding of choice from frontal cortex (MOs, MOp and PL), striatum (CP), and midbrain (MRN, SCm, ZI and SNr) did not significantly differ ($P > 0.05$, two-way ANOVA). Shaded regions, s.e.m. across recordings. **d**, Left, jPECC analysis

related to movements. Regions that showed differences in pre-stimulus firing rate between task and passive contexts also showed similar differences between Go and Miss trials (Fig. 5d). Indeed, it was possible to predict whether a mouse would respond to the stimulus on a given trial from the ‘engagement index’, a projection of pre-stimulus population activity onto the axis defined by the difference of pre-stimulus activity in passive and task contexts (Fig. 5e). This engagement index consistently differed between Go and Miss trials across recordings (Fig. 5f; paired t -test, $P < 10^{-4}$), an effect that could not be fully explained by variability in pupil diameter, in overt movements detectable by video recordings, in the presence of a reward on the previous trial, or in the

shows that population activity in visual cortex predicts that in frontal cortex following a lag of about 40 ms, but only in the period about 200 ms before movement. Right, population activity in midbrain and frontal cortex do not show a consistent lead–lag relationship. **e**, Trial-averaged firing rates of example neurons recorded in the midbrain (top row) and forebrain (bottom row) aligned to contralateral (orange) and ipsilateral choices (blue). **f, g**, Scatter plot of activity of individual midbrain and forebrain neurons at movement onset relative to baseline activity, for trials with contralateral versus ipsilateral choices (estimated from the kernel model). Darker points represent neurons with significant choice encoding. **h**, Summary of **f** and **g** on a brain map. Red and tan indicate regions containing neurons of unilateral or bilateral selectivity. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (version 3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).

inter-trial interval (Extended Data Fig. 9b–k). This index is therefore distinguishable from correlates of movement, reward and arousal, and represents a specific brain-wide neural signature of engagement.

Discussion

Using brain-wide recordings of neuronal populations, we revealed organizing principles of the distribution of neurons carrying distinct correlates of a visual choice behaviour. Neurons non-selectively encoding action are widely distributed. Neurons encoding choice are rarer, are less widely distributed, and exhibit unilateral encoding in midbrain

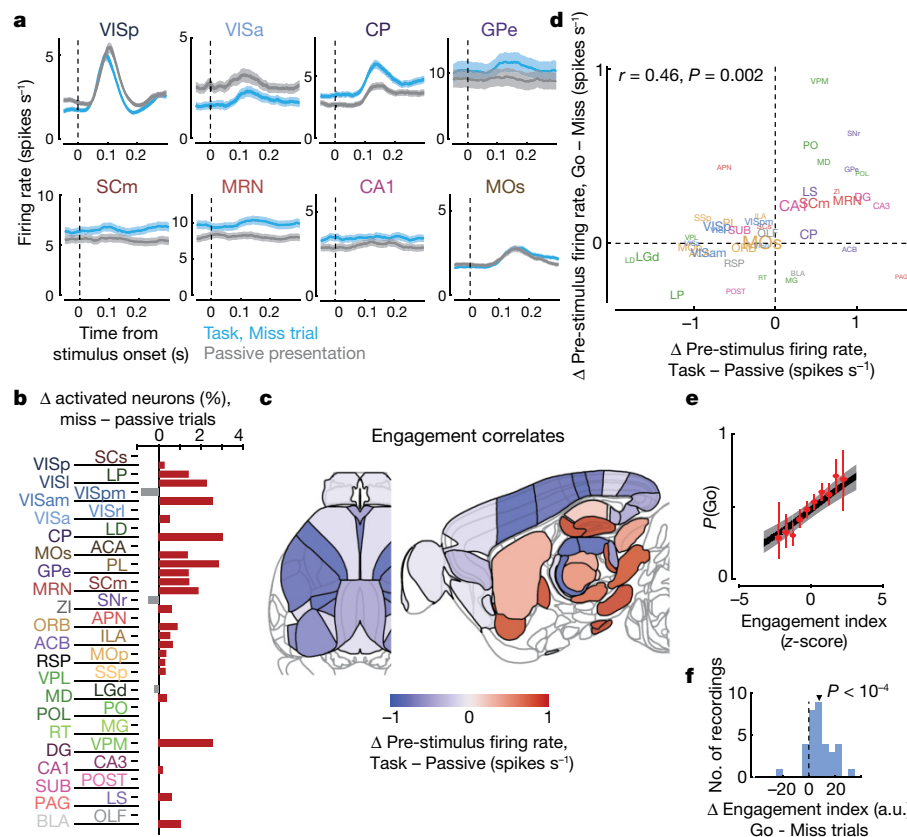


Fig. 5 | Task engagement correlates differently with cortical and subcortical activity. **a**, Comparison of population average spiking activity for several brain regions, for task-context trials when contralateral stimuli were presented but mice did not respond (that is, Miss trials, blue) and for stimulus presentations during the passive context (grey). Visual stimulus contrasts were matched between the two conditions. **b**, Excess fraction of neurons significantly responsive on Miss trials compared to passive condition, for matched contrast stimuli. **c**, Brain map showing difference in pre-stimulus firing rate between task and passive conditions, averaged over all neurons in each region. **d**, Scatter plot showing difference in pre-stimulus rate between task and passive contexts (x axis) and between Go and Miss trials within the task context (y axis), averaged over all neurons in a region. Text size indicates number of neurons in the analysis

(range 130–1,534). **e**, On each trial, an engagement index is computed by projecting pre-stimulus population activity onto a vector defined by the each neuron's rate difference between task and passive contexts. The graph shows probability of Go response as a function of z scored engagement index. Red, movement probability for each bin of engagement index (error bars indicate s.e.m. across trials). Black, logistic regression fit with 95% confidence bands (grey). **f**, Histogram of differences in pre-stimulus engagement index for Go versus Miss trials for each recording. Inverted triangle represents the mean value across recordings (mean = 8.42 AU). AU, arbitrary units. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (version 3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).

and bilateral encoding in forebrain regions. Correlates of engagement are characterized by enhanced subcortical activity and suppressed neocortical activity.

Neurons with action correlates are found globally: neurons in nearly every brain region were non-selectively activated in the moments leading up to movement onset. This global representation of action is consistent with reports of widespread action correlates in multiple species^{14,34,38}, and suggests that non-specific action correlates may in fact be ubiquitous in the mouse brain, cortically and subcortically. These signals may comprise forms of corollary discharge³⁹, but cannot reflect sensory re-afference as they were observed before movement onset. Global non-selective action correlates may underlie brain-wide task-related activity observed in rodents²⁹ and humans²³. This ubiquitous presence of non-selective action correlates underscores the advantage of multi-alternative tasks for revealing the neural correlates of behavioural choice: Go–NoGo tasks cannot distinguish neurons that fire non-specifically for any action from neurons selective for specific choices.

The set of regions encoding choice is spatially restricted and is characterized by qualitatively distinct midbrain and forebrain components. It includes many of the regions classically implicated in choice behaviour including frontal cortex⁵, striatum⁶, substantia nigra pars reticulata⁷ and the deep layers of the superior colliculus^{8,40}, but also

unexpected regions including the midbrain reticular nucleus and zona incerta. These regions also contained neurons encoding visual stimuli, even during passive stimulus presentation; whether visual neurons would be found in these ‘motor’ areas in untrained animals is not clear from these data. Our analyses revealed a striking anatomical organizing principle: choice neurons in the forebrain (neocortex and striatum) are enhanced before both contra- and ipsilateral choices and can prefer either, but choice neurons in the midbrain are almost exclusively enhanced for contralateral choices and often suppressed for ipsilateral choices. Despite this distinct encoding, we could not distinguish the timing of choice-related signals between these regions, an observation parsimoniously explained by a recurrent loop across them (Supplementary Discussion; Supplementary Fig. 4a).

Brain-wide correlates of engagement are characterized by enhanced subcortical activity and suppressed neocortical activity before onset of visual stimulus. Engagement-related cortical suppression might seem surprising given that visual cortex is required for performance of this task^{3,41}. However, increased arousal has been associated with reduced spiking activity and hyperpolarization in multiple cortical areas^{42,43}, an effect that may improve signal-to-noise ratios of sensory representations. Enhanced activity in subcortical areas, by contrast, brings activity in these regions closer to the level at which actions are initiated,

providing a potential mechanism for increased probability of action in engaged states (Supplementary Discussion; Supplementary Fig. 4b–d).

In summary, we have identified organizing principles for the distribution and character of the neuronal correlates of a lateralized visual discrimination task across the mouse brain. Future work will be required to determine the circuit mechanisms that enforce these principles, how they extend to areas such as cerebellum and brainstem omitted from the current survey, and the degree to which similar principles govern the neural correlates of different choice tasks.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1787-x>.

- Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
- Steinmetz, N. A., Koch, C., Harris, K. D. & Carandini, M. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr. Opin. Neurobiol.* **50**, 92–100 (2018).
- Burgess, C. P. et al. High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice. *Cell Rep.* **20**, 2513–2524 (2017).
- Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).
- Romo, R. & de Lafuente, V. Conversion of sensory signals into perceptual decisions. *Prog. Neurobiol.* **103**, 41–75 (2013).
- Ding, L. & Gold, J. I. The basal ganglia's contributions to perceptual decision making. *Neuron* **79**, 640–649 (2013).
- Hikosaka, O., Takikawa, Y. & Kawagoe, R. Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* **80**, 953–978 (2000).
- Felsen, G. & Mainen, Z. F. Midbrain contributions to sensorimotor decision making. *J. Neurophysiol.* **108**, 135–147 (2012).
- Gao, Z. et al. A cortico-cerebellar loop for motor planning. *Nature* **563**, 113–116 (2018).
- Chabrol, F. P., Blot, A. & Mrcic-Flogel, T. D. Cerebellar contribution to preparatory activity in motor neocortex. *Neuron* **103**, 506–519.e4 (2019).
- Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
- Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
- Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y. & Petersen, C. C. H. Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat. Neurosci.* **16**, 1671–1677 (2013).
- Stringer, C. et al. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, eaav7893 (2019).
- Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible sensorimotor decisions. *Science* **348**, 1352–1355 (2015).
- Allen, W. E. et al. Global representations of goal-directed behavior in distinct cell types of mouse neocortex. *Neuron* **94**, 891–907.e6 (2017).
- Poort, J. et al. Learning enhances sensory and multiple non-sensory representations in primary visual cortex. *Neuron* **86**, 1478–1490 (2015).
- Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebri, S. & Movshon, J. A. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13**, 87–100 (1996).
- Yang, H., Kwon, S. E., Severson, K. S. & O'Connor, D. H. Origins of choice-related activity in mouse somatosensory cortex. *Nat. Neurosci.* **19**, 127–134 (2016).
- van Vugt, B. et al. The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* **360**, 537–542 (2018).
- Hernández, A. et al. Decoding a perceptual decision process across cortex. *Neuron* **66**, 300–314 (2010).
- Shuler, M. G. & Bear, M. F. Reward timing in the primary visual cortex. *Science* **311**, 1606–1609 (2006).
- Gonzalez-Castillo, J. et al. Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *Proc. Natl Acad. Sci. USA* **109**, 5487–5492 (2012).
- Vickery, T. J., Chun, M. M. & Lee, D. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* **72**, 166–177 (2011).
- Svoboda, K. & Li, N. Neural mechanisms of movement planning: motor cortex and beyond. *Curr. Opin. Neurobiol.* **49**, 33–41 (2018).
- Jacobs, E. A. K., Steinmetz, N. A., Carandini, M. & Harris, K. D. Cortical state fluctuations during sensory decision making. Preprint at *bioRxiv*, <https://doi.org/10.1101/348193> (2018).
- Mackworth, N. H. The breakdown of vigilance during prolonged visual search. *Q. J. Exp. Psychol.* **1**, 6–21 (1948).
- Makeig, S. & Inlow, M. Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalogr. Clin. Neurophysiol.* **86**, 23–35 (1993).
- Allen, W. E. et al. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* **364**, eaav3932 (2019).
- Raichle, M. E. et al. A default mode of brain function. *Proc. Natl Acad. Sci. USA* **98**, 676–682 (2001).
- Sridharan, D., Steinmetz, N. A., Moore, T. & Knudsen, E. I. Distinguishing bias from sensitivity effects in multialternative detection tasks. *J. Vis.* **14**, 2194077 (2014).
- Pachitariu, M., Steinmetz, N. A., Kadir, S., Carandini, M. & Harris, K. D. Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv. Neural Inf. Process. Syst.* **29**, 1–9 (2016).
- Rossant, C. et al. Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).
- Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
- Aertsen, A. M., Gerstein, G. L., Habib, M. K. & Palm, G. Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *J. Neurophysiol.* **61**, 900–917 (1989).
- Gerstein, G. L. & Perkel, D. H. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* **164**, 828–830 (1969).
- Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259 (2019).
- Ahrens, M. B. et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–477 (2012).
- Crappe, T. B. & Sommer, M. A. Corollary discharge across the animal kingdom. *Nat. Rev. Neurosci.* **9**, 587–600 (2008).
- Horwitz, G. D., Batista, A. P. & Newsome, W. T. Representation of an abstract perceptual decision in macaque superior colliculus. *J. Neurophysiol.* **91**, 2281–2296 (2004).
- Zatka-Haas, P., Steinmetz, N. A., Carandini, M. & Harris, K. D. Distinct contributions of mouse cortical areas to visual discrimination. Preprint at *bioRxiv* <https://doi.org/10.1101/501627> (2018).
- Vinck, M., Batista-Brito, R., Knoblich, U. & Cardin, J. A. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron* **86**, 740–754 (2015).
- Shimaoka, D., Harris, K. D. & Carandini, M. Effects of arousal on mouse sensory cortex depend on modality. *Cell Rep.* **22**, 3160–3167 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986) and under personal and project licenses released by the Home Office following appropriate ethics review.

Mice

Experiments were performed on 10 male and female mice, between 11 and 46 weeks of age (Supplementary Table 1). Multiple genotypes were used, including: Ai95;Vglut1-Cre (B6J.Cg-Gt(ROSA)26Sor^{tm95.1(CAG-GCaMP6f)Hze/MwarJ} crossed with B6;129S-Slc17a7^{tm1.1(Cre)Hze/J}), TetO-G6 s;Camk2a-tTa (B6;DBA-Tg(tetO-GCaMP6 s)2Niell/J crossed with B6.Cg-Tg(Camk2a-tTa)1Mmay/DbolJ), Snap25-G6 s (B6.Cg-Snap25^{tm3.1Hze/J}), Vglut1-Cre, and wild-type (C57Bl6/J). None of these lines are known to exhibit aberrant epileptiform activity⁴⁴. Of 13 mice initially trained for inclusion in this study, three developed health complications before training completed and were not recorded. The other 10 successfully learned the task (see criteria below) and were included. The sample sizes ($n = 10$ mice; $n = 39$ recording sessions; $n = 29,134$ neurons) were not determined with a power analysis.

Surgery

A brief (around 1 h) initial surgery was performed under isoflurane (1–3% in O₂) anaesthesia to implant a steel headplate (approximately 15 × 3 × 0.5 mm, 1 g) and, in most cases, a 3D-printed recording chamber. The chamber was a semi-conical, opaque piece of polylactic acid with a 12 mm diameter upper surface and a lower surface designed to fit to the shape of an average mouse skull, exposing approximately the area from 3.5 mm anterior to 5.5 mm posterior to bregma, and 4.5 mm left to 4.5 mm right, and narrowing near the eyes. The implantation method largely followed the method of Guo et al.⁴⁵ with some modifications and was previously described⁴⁴. In brief, the dorsal surface of the skull was cleared of skin and periosteum and prepared with a brief application of green activator (Super-Bond C&B, Sun Medical). The chamber was attached to the skull with cyanoacrylate (VetBond; World Precision Instruments) and the gaps between the cone and the skull were filled with L-type radiopaque polymer (Super-Bond C&B). A thin layer of cyanoacrylate was applied to the skull inside the cone and allowed to dry. Thin layers of UV-curing optical glue (Norland Optical Adhesives #81, Norland Products) were applied inside the cone and cured until the exposed skull was covered. The head plate was attached to the skull over the interparietal bone with Super-Bond polymer, and more polymer was applied around the headplate and cone.

After recovery, mice were given three days to recover while being treated with carprofen, then acclimated to handling and head-fixation before training.

Two-alternative unforced choice task

The two-alternative unforced choice task design was described previously³. Mice sat on a plastic apparatus with forepaws on a rotating wheel surrounded by three computer screens (Adafruit, LP097QX1) at right angles covering 270 × 70 degrees of visual angle (d.v.a.). Each screen was roughly 11 cm from the mouse's eyes at its nearest point and refreshed at 60 Hz. The screens were fitted with Fresnel lenses (Wuxi Bohai Optics, BHPA220-2-5) to ameliorate reductions in luminance and contrast at larger viewing angles near their edges, and these lenses were coated with scattering window film ('frostbite', The Window Film Company) to reduce reflections. The wheel was a ridged rubber Lego wheel affixed to a rotary encoder (Kubler 05.2400.1122.0360). A plastic tube for delivery of water rewards was placed near the subject's mouth. Licking behaviour was monitored by attaching a piezo film (TE Connectivity, CAT-PFS0004) to the plastic tube and recording its voltage. Experiments were run with Rigbox software for MATLAB (Mathworks, Inc.)⁴⁶. Full details of the experimental apparatus

including detailed parts list can be found at <http://www.ucl.ac.uk/cortexlab/tools/wheel>.

A trial was initiated after the subject had held the wheel still for a short interval (duration uniformly distributed between 0.2–0.5 s on each trial; Fig. 1b). At trial initiation, visual stimuli were presented at the centre of the left and right screens, or directly left and right of the subject. These stimulus locations were in the central of the monocular zones of the mouse's visual field so that no eye or head movements were required for the mice to see them. The stimulus was a Gabor patch with orientation 45°, sigma 9 d.v.a., and spatial frequency 0.1 cycles per degree. After stimulus onset there was a random delay interval of 0.5–1.2 s, during which time the subject could turn the wheel without penalty, but visual stimuli were locked in place and rewards could not be earned. The mice nevertheless typically responded immediately to the stimulus onset. At the end of the delay interval, an auditory tone cue was delivered (8 kHz pure tone for 0.2 s) after which the visual stimulus position became coupled to movements of the wheel. Wheel turns in which the top surface of the wheel was moved to the subject's right led to rightward movements of stimuli on the screen, that is, a stimulus on the subject's left moved towards the central screen. Put another way, clockwise turns of the wheel, from the perspective of the mouse, led to clockwise movement of the stimuli around the subject. A left or right turn was registered when the wheel was turned by an amount sufficient to move the visual stimuli by 90 d.v.a. in either direction (approximately 20 mm of movement of the surface of the wheel). When at least one stimulus was presented, the subject was rewarded for driving the higher contrast visual stimulus to the central screen (if both stimuli had equal contrast, left and right turns were rewarded with 50% probability). When no stimuli were presented, the subject was rewarded if no turn was registered during the 1.5 s following the Go cue. There were therefore three trial outcomes that could lead to reward depending on the stimulus condition (left turn, right turn, no turn), and in this sense the task was a three-alternative task. Immediately following registration of a choice or expiry of the 1.5 s window, feedback was delivered. If correct, feedback was a water reward (2–3 µl) delivered by the opening of a valve on the water tube for a calibrated duration. If incorrect, feedback was a white noise sound played for 1 s. During the 1 s feedback period, the visual stimulus remained on the screen. After a subsequent inter-trial interval of 1 s, the mouse could initiate another trial by again holding the wheel still for the prescribed duration.

Trials of different contrast conditions were randomly interleaved. The experimenter was not blinded to contrast condition either during data acquisition or during analysis.

Training protocol

Mice were trained on this task with the following shaping protocol. First, high-contrast stimuli (50 or 100%) were presented only on the left or the right, with an unlimited choice window, and repeating trial conditions following incorrect choices ('repeat on incorrect'). Once mice achieved high accuracy and initiated movements rapidly—approximately 70 or 80% performance on non-repeat trials, and with reaction times nearly all <1 s, but at the experimenter's discretion—trials with no stimuli were introduced, again repeating on incorrect. Once subjects responded accurately on these trials (70 or 80% performance, at experimenter's discretion), lower contrast trials were introduced without repeat on incorrect. Finally, contrast comparison trials were introduced, starting with high versus low contrast, then high versus medium and medium versus low, then trials with equal contrast on both sides. The final proportion of trials presented was weighted towards easy trials (high contrast versus zero, high versus low, medium versus zero, and no-stimulus trials) to encourage high overall reward rates and sustained motivation.

On most trials for which subjects eventually made a left or right turn by the end of the trial, the subjects responded immediately to the stimulus presentation, turning the wheel within 400 ms of stimulus

appearance ($64.9 \pm 14.0\%$ (mean \pm s.d.), $n = 39$ sessions), nearly always in the same direction as their final choice ($96.6 \pm 3.4\%$). For this study, data analyses focused on this initial 400 ms period, and we defined left and right choice trials as those in which this period contained the onset of a clockwise or counterclockwise turn of sufficient amplitude (90 d.v.a.), and NoGo trials as those in which it contained no detectable movement. To exclude trials in which wheel turns were coincidentally made before subjects could respond to the stimuli, only trials with movement onset between 125 to 400 ms post-stimulus onset, or with no movement of any kind during the window from -50 to 400 ms post-stimulus onset, were included. Trials with other movements, that were detectable but would not have resulted in registering a choice by the end of the movement, were excluded.

The algorithm for detecting wheel movement onsets ('find-WheelMoves3', <https://github.com/cortex-lab/wheelAnalysis/blob/master/+wheel/findWheelMoves3.m>) was designed to identify the earliest moment at which the wheel began detectably moving. First, non-movement periods were identified as those which had less than 1.1 mm of wheel movement over 0.2 s duration. Next, a morphological closure (dilation then erosion) of these periods was performed with size 0.1 s to remove gaps smaller than this. Finally, the timing of the ends of the non-movement periods were refined by looking sequentially backwards in time to identify the first moment at which the position deviated by more than a smaller threshold of 0.2 mm. The double-threshold procedure (first 1.1 mm, then looking backwards for 0.2 mm) was necessary because 0.2 mm is just two units of the rotary encoder's measurement, and these two-unit steps could happen owing to noise at any time. In this way, the movement onsets (and consequently the reaction times) were measured at the resolution of the rotary encoder. Smaller detection thresholds would lead to earlier detection of wheel turns, but potentially at the risk of false-positive detections. To assess how our detector performed, we decoded the subject's choice from the instantaneous wheel velocity (difference in wheel position between t and $t-10$ ms) at different times relative to detected movement onset (Extended Data Fig. 1q). The decoder performed essentially at chance 50 ms before movement onset (47.7%) compared to near perfect performance 50 ms after movement onset (94.6%).

Behavioural trials when the mouse was disengaged were excluded from analysis. These trials were defined as Miss trials (stimulus present but wheel not turned) preceded by two or more other Miss trials, as well as all NoGo trials occurring consecutively at the end of the session.

When analysing activity following reward delivery (Extended Data Fig. 4e), only correct NoGo trials were included, that is, trials with no visual stimulus and no wheel movement.

Sessions were included when at least 12 trials of each type (left, right and NoGo) could be included for analysis, and when anatomical localization was sufficiently confident (see below).

For analyses requiring matching stimulus contrasts across trials with different choices, we considered all trials with contralateral stimulus contrast greater than zero, and split them by low, medium, and high contralateral contrast. For each contrast level, we counted the number of trials with that contrast and each response type (left or right; NoGo; or passive condition). We took the minimum of these three numbers, and selected that many trials randomly from each group. This resulted in three sets of trials—trials with left or right choices; trials with NoGos; and trials in the passive condition—which each contained low, medium and high contralateral contrasts but which all contained exactly the same numbers of each contrast. When fewer than 10 such trials could be found, the session was excluded for the matched-contrast analyses ($n = 34$ of 39 sessions included).

Video monitoring

Eye and body movements were monitored by illuminating the subject with infrared light (830 nm, Mightex SLS-0208-A). The right eye was monitored with a camera (The Imaging Source, DCM 23U618) fitted

with zoom lens (Thorlabs MVL7000) and long-pass filter (Thorlabs FEL0750), recording at 100 Hz. Body movements were monitored with another camera (same model but with a different lens, Thorlabs MVL16M23) situated above the central screen, recording at 40 Hz.

Neuronal recordings

Recordings were made using Neuropixels (Phase3A Option 3) electrode arrays¹, which have 384 selectable recording sites out of 960 sites on a 1-cm shank. Probes were mounted to a custom 3D-printed polylactic acid piece and affixed to a steel rod held by a micromanipulator (uMP-4, Sensapex Inc.). To allow later track localization, before insertion probes were coated with a solution of DiI (ThermoFisher Vybrant V22888 or V22885) by holding 2 μ l in a droplet on the end of a micropipette and touching the droplet to the probe shank, letting it dry, and repeating until the droplet was gone, after which the probe appeared pink.

On the day of recording or within two days before, mice were briefly anaesthetized with isoflurane while one or more craniotomies were made, either with a dental drill or a biopsy punch. After at least three hours of recovery, mice were head-fixed in the setup. Probes had a soldered connection to short external reference to ground; the ground connection at the headstage was subsequently connected to an Ag/AgCl wire positioned on the skull. The craniotomies as well as the wire were covered with saline-based agar. The agar was covered with silicone oil to prevent drying. In some experiments, a saline bath was used rather than agar. Two or three probes were advanced through the agar and through the dura, then lowered to their final position at approximately $10 \mu\text{m s}^{-1}$. Electrodes were allowed to settle for around 15 min before starting recording. Recordings were made in external reference mode with local field potential gain = 250 and action potential gain = 500. Recordings were repeated at different locations on each of multiple subsequent days (Supplementary Table 2), performing new craniotomy procedures as necessary. All recordings were made in the left hemisphere. The ability of a single probe to record from multiple areas, and the use of multiple probes simultaneously, led to a number of areas being recorded simultaneously in each session (Supplementary Table 3).

Passive stimulus presentation

After each behaviour session, we performed a passive replay experiment while continuing to record from the same electrodes. Mice were presented with two types of sensory stimuli without possibility of receiving reward for any behaviour: replay of task stimuli; and sparse flashed visual stimuli for receptive field mapping.

The replayed task stimuli were: left and right visual stimuli of each contrast; some combinations of left and right visual stimuli simultaneously; Go cue beeps; white noise bursts; and reward valve clicks (but with a manual valve closed so that no water was delivered). These stimuli were replayed at 1–2 s randomized intervals for 10 or 25 randomly interleaved repetitions each.

Receptive fields were mapped with white squares of 8 d.v.a. edge length, positioned on a 10×36 grid (some stimulus positions were located partially off-screen) on a black background. The stimuli were shown for 10 monitor frames (167 ms) at a time, and their times of appearance were independently randomly selected to yield an average rate of approximately 0.12 Hz.

Data analysis

The data were automatically spike sorted with Kilosort³² (<https://github.com/cortex-lab/Kilosort>) and then manually curated with the phy GUI (<https://github.com/kwikteam/phy>). Extracellular voltage traces were preprocessed with common-average referencing⁴⁷: subtracting each channel's median to remove baseline offsets, then subtracting the median across all channels at each time point to remove artefacts. During manual curation, each set of events (unit) detected by a particular template was inspected and if the events comprising the unit were judged to correspond to noise (zero or near-zero amplitude;

non-physiological waveform shape or pattern of activity across channels) rather than spikes, the entire unit was discarded. Units containing low-amplitude spikes, spikes with inconsistent waveform shapes, and/or refractory period contamination were labelled as ‘multi-unit activity’ and not included for further analysis. Finally, each unit was compared to similar, spatially neighbouring units to determine whether they should be merged, on the basis of spike waveform similarity, drift patterns or cross-correlogram features. Units were also excluded if their average rate in the analysis window (stimulus onset to 0.4 s after; trial firing rate) was less than 0.1 Hz. Units passing these criteria were considered to reflect the spiking activity of a neuron.

Neurons were only included for further analysis when at least 13 neurons passing the above criteria were identified as coming from the same brain region, in the same experiment. Furthermore, brain regions were only included for which recordings from at least two subjects had sufficient numbers of neurons.

To determine whether a neuron’s firing rate was significantly modulated during the task (Supplementary Fig. 1), a set of six statistical tests were used to detect changes in activity during various task epochs and conditions: (1) Wilcoxon sign-rank test between trial firing rate (rate of spikes between stimulus onset and 400 ms post-stimulus) and baseline rate (defined in period -0.2 to 0 s relative to stimulus onset on each trial); (2) sign-rank test between stimulus-driven rate (firing rate between 0.05 and 0.15 s after stimulus onset) and baseline rate; (3) sign-rank test between pre-movement rates (-0.1 to 0.05 s relative to movement onset) and baseline rate (for trials with movements); (4) Wilcoxon rank-sum test between pre-movement rates on left choice trials and those on right choice trials; (5) sign-rank test between post-movement rates (-0.05 to 0.2 s relative to movement onset) and baseline rate; (6) rank-sum test between post-reward rates (0 to 0.15 s relative to reward delivery for correct NoGos) and baseline rates. A neuron was considered active during the task, or to have detectable modulation during some part of the task, if any of the P values on these tests were below a Bonferroni-corrected alpha value ($0.05/6 = 0.0083$). However, because the tests were coarse and would be relatively insensitive to neurons with transient activity, a looser threshold was used to determine the neurons included for statistical analyses (Figs. 3–5): if any of the first four tests (that is, those concerning the period between stimulus onset and movement onset) had a P value less than 0.05 .

In determining the neurons statistically significantly responding during different task conditions (Figs. 2d–h, right sub-panels, 5b), the mean firing rate in the post-stimulus window (0 to 0.25 s), taken across trials of the desired condition, was z-scored relative to trial-by-trial baseline rates (from the window -0.1 to 0) and taken as significant when this value was >4 or <-4 , equivalent to a two-sided t -test at $P < 10^{-4}$.

For visualizing firing rates (Extended Data Fig. 4), the activity of each neuron was then binned at 0.005 s, smoothed with a causal half-Gaussian filter with standard deviation 0.02 s, averaged across trials, smoothed with another causal half-Gaussian filter with standard deviation 0.03 s, baseline subtracted (baseline period -0.02 to 0 s relative to stimulus onset, including all trials in the task), and divided by baseline $+ 0.5$ spikes s^{-1} . Neurons were selected for display if they had a significant difference between firing rates on trials with both stimuli and movements versus trials with neither, using a sliding window 0.1 s wide and in steps of 0.005 s (rank-sum $P < 0.0001$ for at least three consecutive bins).

Visual receptive fields (Extended Data Fig. 2d) were determined by sparse noise mapping outside the context of the behavioural task. The evoked rates for each presentation were measured as the spike count in the 200 ms following stimulus onset. The rates evoked by stimuli at the peak location and surrounding four nearest locations were combined and compared to the rates for all locations >45 d.v.a. from the peak location using a Wilcoxon rank-sum test. Any neurons for which the p value of the test was less than 10^{-6} were counted as having a significant visual receptive field. Note that neurons are included in analyses regardless

of receptive field location; in particular, recorded LGd neurons did not have receptive field locations overlapping with task stimuli.

Kernel regression analysis

To identify choice-selective neurons, we began by fitting a kernel-regression model^{48–50}. In this analysis, the firing rate of each neuron is described as a linear sum of temporal filters aligned to task events. For the current study, only visual stimulus onset and wheel movement onset kernels were required, since we consider here only the period in between the two. In the model, the predicted firing rate $f_n(t)$ for neuron n is given as

$$f_n(t) = \sum_c \sum_{t_s \in S_c} K_{c,n}(t - t_s) + \sum_{t_m \in M} (K_{m,n}(t - t_m) + D_m K_{D,n}(t - t_m))$$

Here, c represents the six stimulus types (contralateral low, medium, or high, or ipsilateral low, medium, or high), S_c represents the set of times for which this contrast appeared, and $K_{c,n}(t)$ represents the Vision kernel function of this contrast for neuron n . M represents the set of movement times and $K_{m,n}(t)$ represents the Action kernel for neuron n ; D_m represents direction of movement m (encoded as ± 1), and $K_{D,n}$ represents the Choice kernel for neuron n . The Vision kernels $K_{c,n}(t)$ are supported over the window -0.05 to 0.4 s relative to stimulus onset, and the Action and Choice kernels are supported over the window -0.25 to 0.025 s relative to movement onset. Before estimating the kernels, the discretized firing rates $f_n(t)$ for each neuron were estimated by binning spikes into 0.005 -s bins and then smoothing with a causal half-Gaussian filter with standard deviation 0.025 s. The Vision kernels therefore contain $L_c = 90$ time bins, whereas Action and Choice kernels contain $L_d = 55$ time bins.

The large number of parameters to be fit, combined with the relatively small number of trials of each type pose a challenge for estimation. We devised a solution to this problem that makes use of the large number of neurons recorded using reduced-rank regression (Extended Data Fig. 5b), which we found to give better cross-validated results (see next section).

First, for each kernel to be fit, we construct a Toeplitz predictor matrix (Extended Data Fig. 5d). For stimuli of contrast c , we define a Toeplitz predictor matrix \mathbf{P}_c of size $T \times L_c$, in which T is the total number of time points in the training set, and L_c is the number of lags required for the Vision kernels. The predictor matrix contains diagonal stripes starting each time a visual stimulus of contrast c is presented: $\mathbf{P}_c(t, i) = 1$ if $t - i \in S_c$ and 0 otherwise. Predictor matrices of size $T \times L_d$ were defined similarly for the Action and Choice kernels, and the six stimulus predictor matrices and two movement predictors are horizontally concatenated to yield a global prediction matrix \mathbf{P} of size $T \times 650$. The total length of all kernels for one neuron is $650 = 6L_c + 2L_d$.

The simplest approach to fit the kernel shapes would be to minimize the squared error between true and predicted firing rate using linear regression. To do this, we would horizontally concatenate the rate vectors of all N neurons together into a $T \times N$ matrix \mathbf{F} , and estimate the kernels for each neuron by finding a matrix \mathbf{K} of size $650 \times N$ to minimize the squared error $E = \|\mathbf{F} - \mathbf{PK}\|^2$, using the total sum square (Frobenius) norm. However, as each \mathbf{k}_n has 650 parameters, linear regression results in noisy and overfit kernels when fit to a single neuron, particularly given the high trial-to-trial variability of neuronal firing. Although expressing the kernels as a sum of basis functions can reduce the number of required parameters⁴⁸, the success of this method depends strongly on the choice of basis functions, and the appropriate choice will differ depending on properties of the task and stimuli. The large number of neurons in the current dataset enables an alternative approach.

This approach is based on reduced-rank regression⁵¹, which allows regularized estimation by factorizing the kernel matrix \mathbf{K} into the product of a $650 \times r$ matrix \mathbf{B} and a $r \times N$ matrix \mathbf{W} , minimizing the total

error: $E = \|\mathbf{F} - \mathbf{PBW}\|^2$. The $T \times r$ matrix \mathbf{PB} may be considered as a set of temporal basis functions, which can be linearly combined to estimate each neuron's firing rate over the whole training set. Reduced rank regression ensures that these basis functions are ordered, so that predicting population activity from only the first r columns will result in the best possible prediction from any rank r matrix.

To estimate each neuron's kernel functions, we estimated a weight vector \mathbf{w}_n to minimize an error $E_n = \|\mathbf{f}_n - \mathbf{PBw}_n\|^2$ for each neuron with elastic net regularization (using the package `cvglmnet` for Matlab (http://www.stanford.edu/~hastie/glmnet_matlab/) with parameters $\alpha = 0.5$ and $\lambda = 0.5$), and used cross-validation to determine the optimal number of columns r_n of \mathbf{PB} to keep when predicting neuron n . The kernel functions for neuron n were then unpacked from the 650-dimensional vector obtained by multiplying the first r_n columns of \mathbf{B} by \mathbf{w}_n . Neurons with total cross-validated variance explained of $<2\%$ were excluded from analyses.

Comparison of reduced-rank model to alternative models

To demonstrate the validity of this reduced-rank kernel method, we compared its performance of the reduced-rank regression method to two alternative approaches to spike train prediction: (1) fitting regression to the Toeplitz predictor matrix directly, and (2) a model with raised cosine basis functions (Extended Data Fig. 5d).

The Toeplitz predictor matrix was the matrix \mathbf{P} described above. The cosine predictor matrix was constructed similarly, but with each row containing a raised cosine of width 100 ms, and spaced by 25 ms. The cosine predictor matrix therefore contained $L_c = 18$ rows for each of the 6 contrasts, and $L_d = 11$ time rows for each movement kernel, for 130 predictors in total (Extended Data Fig. 5d).

To evaluate the performance of these three methods, we fit regression weights with elastic net regression as described above, and evaluated performance as the percentage of variance explained (fivefold cross-validation across trials). To compare models fairly, the number of columns included in the reduced-rank model was not allowed to vary per neuron as described above, but was instead fixed at $n = 18$ components. The reduced-rank projection matrix \mathbf{B} was not itself cross-validated. The reduced-rank method outperformed both other methods (Extended Data Fig. 5e)

To estimate the degree to which differing performance of the models arose from over- versus under-fitting, we computed the 'proportion of overfit explained variance' as

$$\frac{CV_{\text{train}} - CV_{\text{test}}}{CV_{\text{train}}}$$

where CV_{train} is the training-set variance explained and CV_{test} is the test-set variance explained, thus quantifying the difference between the two (a measure of overfitting) relative to the total variance explained (Extended Data Fig. 5f).

Determining individual neuron selectivity

To assess the selectivity of individual neurons for each kernel, we used a nested approach. We first fit the activity of each neuron using the reduced-rank regression procedure above (including deriving a new basis set), but excluding the kernel to be tested. We subtracted this prediction from the raw data to yield residuals, representing aspects of the neuron's activity not explainable from the other kernels. We then repeated the reduced-rank regression procedure one more time, using the residual firing rates as the independent variable, and using only the test kernel. The cross-validated quality of this fit determined the variance explainable only by the test kernel. If this variance explained was $>2\%$, the neuron was deemed selective for that kernel and was included in Fig. 3d,e or Fig. 4b.

In principle, inaccuracies in the model fits from one kernel could leave variability to be explained by another correlated variable, resulting in

false positives from this test. For instance, since motor actions are correlated with visual stimuli, activity related to task-unrelated movements (for example, as reported in ref.¹⁷) could appear to be visually related activity if it was not accurately captured by the Action and Choice kernels. However, we consider it unlikely that a significant proportion of the contralateral vision correlates reported here arise from such a confound, since the same argument should apply to ipsilateral vision correlates, and we found very little of such correlates (Extended Data Fig. 5g).

Another potential source of error is that neurons were chosen for inclusion in the analysis based on criteria that could relate to the results of the analysis. Specifically, one of the criteria by which a neuron could be included was the observation of significant differences between left and right choice trials (see above). However, our empirically estimated false discovery rate is very low (0.3%, Extended Data Fig. 5h), and it should not differ between brain regions. Thus, the observation that choice-selective neurons were not found in most brain regions studied is an internal control, showing that false discovery of neurons based on analysis inclusion criteria cannot account for our findings.

Estimation of false-positive rate for determining individual neuron selectivity with reduced-rank kernel regression

To choose the threshold for counting a neuron as selective, we searched for a value giving low false-positive error rates for choice selective neurons. To estimate false-positive rates, we performed a shuffle analysis, relabelling each trial with a left or right choice with a randomly drawn choice from another left or right choice trial, without replacement. The analysis was then repeated from the start, including fitting the reduced-rank regression and the cross-validated nested model. We selected the threshold for counting a neuron as choice selective to ensure a low false-positive error rate as assessed with this measure (0.33%; Extended Data Fig. 5h).

Population decoding of task correlates

To perform population decoding (Fig. 4c; Extended Data Fig. 5g), we began with the residual firing rates produced as described above, produced by fitting without a test kernel. We then split trials in a binary fashion: trials that had versus those that did not have an ipsilateral stimulus; those having versus not having a contralateral stimulus; those having versus not having any movement (either left or right); those having left choice versus having right choice (considering only trials with one of the two). We identified a population coding direction encoding the difference between the two sets of trials, by fitting an L1-regularized logistic regression on data from training trials, using the period 0.05 to 0.15 s relative to stimulus onset for Vision decoding, and the period -0.05 to 0 s relative to movement onset for Action or Choice decoding. We then predicted the binary category of test data by projecting firing rates from test set trials, from each time point during the trial, onto the weight vector of the logistic regression. Although it is in principle possible that these signals are encoded in a nonlinearly separable way, the robust predictions obtained suggest information can be read out linearly. The population decoding was taken as the difference between projections between test-set trials of each binary category. For Action decoding, where trials with left or right choices were compared to those with neither, a 'movement onset time' was chosen for trials without a movement randomly from the distribution of movement onset times on left and right choice trials.

To statistically compare decoding time course across areas, we took the population decoding score from each key area in each recording ($n = 29$ populations from frontal cortex including MOs, PL and MOp; $n = 29$ populations from midbrain including SCm, MRN, SNr and ZI; $n = 5$ striatum, CP), and normalized each so the mean across recordings within an area was 1 at choice time. We then performed a two-way ANOVA, with factors time relative to movement onset and area (frontal, midbrain and striatum). We found a significant effect of time on choice

decoding (50 d.f., $F = 10.43$, $P < 10^{-71}$) but no significant effect of area (2 d.f., $F = 0.28$, $P > 0.05$) and no significant interaction between time and area (100 d.f., $F = 0.12$, $P > 0.05$).

jPECC analysis

To perform jPECC analysis (Fig. 4d, Extended Data Fig. 8), we took spike counts of individual neurons from simultaneously recorded regions, assembling neurons from all regions within a group into one population vector per region group. The three region groups were: VIS: VISp, VISpm, VISl, VISr, VISa, VISam; midbrain: SCm, MRN, ZI, SNr; and frontal: MOs, MOp, PL. One jPECC per pair of region groups was computed per recording. Spikes were counted in 10-ms bins and smoothed with a half-Gaussian causal filter with 25-ms standard deviation, and normalized by dividing by baseline $+1 \text{ spike s}^{-1}$. Principal components analysis was then performed for each region group, across time points and trials to reduce population activity to ten dimensions. Trials were divided tenfold into training and test sets. Canonical correlation analysis was performed on the training set PCs from each region group, L2 regularized using $\lambda = 0.5$. The test-set PCs were projected onto the top canonical dimension, and the Pearson correlation coefficient was computed between these projections across test-set trials. This process was repeated for each pair of time bins, creating a matrix of cross-validated correlation coefficients, with one entry per pair of time points relative to the event. When representing a single recording's jPECC analysis, the statistical significance of these correlation coefficients was used to grey out non-significant regions (Extended Data Fig. 8b) but this value was not used in further analyses.

To quantify lead-lag relationships across recordings, an asymmetry index was computed by diagonally slicing the jPECC matrix from -50 to $+50$ ms relative to each time point. The average correlation coefficient across the left half of this slice (that is, the average along a vector from $[t-50, t+50]$ to $[t, t]$) was subtracted from the right half of this slice (from $[t, t]$ to $[t+50, t-50]$) to yield the asymmetry index for time point t . This index was computed for each time point t relative to events and the values across recordings were compared to 0 with a t -test.

Engagement index and pre-stimulus analyses

To statistically compare pre-trial firing rates between the task and passive conditions (that is, between trials of active task performance, versus later passive stimulus replay, Extended Data Fig. 9a), we performed a nested multiple ANOVA test, in order to account for correlated variability between neurons within recording sessions. Each observation was a neuron's average measured pre-trial firing rate in the window between 250 and 50 ms before stimulus onset, log transformed ($\log_{10}(x + 1 \text{ spike s}^{-1})$) to make distributions approximately normal. Any trials with detectable wheel movement in this interval were excluded. The ANOVA had three factors: active or passive condition, recording session, and neuron identity (nested within recording session). A separate ANOVA was performed for each brain region, and the null hypothesis of no difference between baseline rates in active and passive conditions for neurons from a given brain region was rejected if the P -value for the active/passive condition factor was less than 0.0012, that is, less than 0.05 after applying a Bonferroni correction for the 42 brain regions tested.

To compute the trial-by-trial 'engagement index', we took the difference in pre-stimulus firing between the average of all task ('active') trials a and of all passive trials p , over the 200-ms period before stimulus onset:

$$x_n = \bar{f}_n^a(-0.2 < t < 0) - \bar{f}_n^p(-0.2 < t < 0)$$

This quantity was computed for each neuron in one of the areas with significant differences between task and passive determined by nested ANOVA analysis (319 ± 32.5 (mean \pm s.e.m.) neurons included per session, $n = 34$ sessions), accumulated into a vector \mathbf{x} and normalized

to unit L2 magnitude for each session. To compute the engagement index for each trial i we computed the dot product $\mathbf{x} \cdot \mathbf{f}$, where \mathbf{f} is the vector of pre-stimulus firing rates for each trial i . For the summary analysis in Fig. 5f, we then computed the mean across Go trials and the mean across Miss trials, and took the difference of the two. The Go and Miss trials included in this analysis were matched for contrast so that difference in visual drive could not influence the difference between trial types. To do this, N trials were selected from each contralateral contrast condition, where N was the minimum of the number of trials at that contrast condition having a Go outcome and the number having a NoGo outcome.

Measurement of pupil area and video motion energy

We measured pupil area (Extended Data Fig. 9g, i, k) from the high-zoom videos of the subject's eye, using DeepLabCut⁵². In approximately 200 training frames randomly sampled across all sessions, 4 points spaced at 90° around the pupil were manually identified, and the network was trained with default parameters. Then, around 100 more frames were manually annotated focusing on frames with errors, and the network re-trained. The pupil area was taken to be the area of an ellipse with major and minor axis lengths given by the distances between opposite pairs of detected points. Some recordings were excluded from these analyses owing to video quality that was unusable for sufficiently accurate measurement of the pupil ($n = 5$ out of 39), primarily owing to obscured pupils due to eyelashes or eyelids.

We measured video motion energy (Extended Data Fig. 9h, i, k) from the low-zoom videos of the frontal aspect of the subjects, which included the face, arms, and part of the torso of the mice. As nothing in the frame except the mouse could move, we calculated total motion energy of the pixels of the video as an index of overt movements of the mouse. This was computed as:

$$\sum_{\text{pixels}} \text{abs}(i_t - i_{t-1})$$

where i_t is the intensity of the pixel on frame t .

Both pupil area and video motion energy were z-scored before generalized linear model (GLM) fitting.

GLM prediction of $P(\text{Go})$ from pre-stimulus variables

To determine the impact of arousal-, reward- and history-related variables on the ability to predict whether the upcoming trial's outcome would be a Go response (that is, a left or right choice), and whether these factors could account for the relationship between Engagement Index and Go/Miss trials, we fit a GLM model to several measures of overt behaviour (Extended Data Fig. 9). These behavioural variables were: inter-trial interval; previous trial reward outcome (coded as 0 or 1); pupil area in the pre-stimulus window (z-scored); and motion energy in the pre-stimulus window (z-scored). A GLM with binomial link function was fit to these variables (Matlab function 'fitglm') to predict whether the following trial had a Go or NoGo outcome. Squared terms were included for pupil area and video motion energy after it was observed that the empirical relationship between each of these and the $P(\text{Go})$ had an inverted U-shape⁵³. Trials were selected for the model fitting to have matched contrast between these two types (see Engagement index above). A deviance test was used to compare this model with a model that additionally had engagement index included as a predictor. A significant value of this test does not indicate that the engagement index suffices to predict $P(\text{Go})$, with other variables making no further contribution; rather, it indicates that the other variables did not fully predict $P(\text{Go})$, and that engagement index can improve this prediction. The population vector analysis in Extended Data Fig. 9k further argues that engagement index relates to $P(\text{Go})$ more closely than does the population vector of these other variables.

Combined-conditions choice probability and detect probability analysis

Choice probability is a non-parametric measure of the difference in firing rate between trials with identical stimulus conditions but different choices. Typically, it is calculated separately for each stimulus condition, but here we used an algorithm that combines observations across stimulus conditions into one number, allowing it to be calculated even for our small number of trials per condition (since there are 16 stimulus conditions, and trials with NoGo responses are excluded for the choice probability calculation). Choice probability is classically calculated as the area under a receiver operating characteristic (ROC) curve, which is equivalent to a Mann–Whitney U statistic, that is, to the probability that a firing rate observation from the trials with one choice is greater than that from trials with the other choice. Accordingly, this can be calculated by comparing each trial of one condition to each of the other condition, counting the number of such comparisons for which the first condition wins, and dividing by the total number of comparisons. To extend the method to a situation of many stimulus conditions but few trials of each condition, we add the numerators and denominators of this ratio across all conditions, and then divide. In this way, the core logic of choice probability—that trials of one choice are only compared to trials of the other choice under identical stimulus conditions—is preserved, but all stimulus conditions can be combined into one number per neuron. This number quantifies the same thing as the classic choice probability, namely, the probability that the spike count of a neuron will be greater for trials of one choice than another, given matched stimulus conditions. To estimate statistical significance, we used a shuffle test in which trial labels (as left or right choice) were randomly permuted within each stimulus condition 2,000 times, and the choice probability was computed for each shuffle. Because this algorithm combines trials from multiple stimulus conditions into a single statistic, we refer to it as the combined-conditions choice probability (ccCP).

We used the same algorithm to compute a detect probability (DP) comparing trials with NoGo outcomes to those with either left or right choices.

Importantly, not all trials can be included in the ccCP analysis. Specifically, the trials from any stimulus condition in which the subject only made left or only made right choices cannot contribute to the ccCP. This is unlike the kernel regression analysis, in which those trials would still contribute to the estimates of the vision and choice kernels (and note that the kernel analysis additionally makes use of reaction time variability to separate these representations). As a result, six sessions had to be excluded for having fewer than ten trials include-able in the ccCP analysis. Notably, two of these six sessions contained some of the MOs neurons with significant choice representations under the kernel analysis, so that 20.8% of MOs choice-selective neurons were not included here.

Focal index

To statistically test the degree to which neurons encoding different task variables were localized, we used a focal index, defined as:

$$F = \frac{\sum (p_a^2)}{(\sum p_a)^2}$$

where p_a is the proportion of neurons in an area selective for the task correlate of interest (action, contralateral vision or choice), as assessed by reduced-rank regression analysis (Fig. 3c, e, 4b). This measure is an adaptation of the sparseness measure of Treves and Rolls³⁴, and would take the value 1 if all neurons were located in a single region, and the value $1/N$ if neurons were equally probable in N regions. Ninety-five per cent confidence intervals were computed using the bootstrap, with a normal approximated interval with bootstrapped bias and standard error (function `bootci` in Matlab).

Anatomical targeting

To select probe insertion trajectories, we first identified desired recording sites, and then designed appropriate trajectories to reach them using the `allen_ccf_npx` gui (A.J. Peters, www.github.com/cortex-lab/allenCCF). In doing so, Allen CCF coordinate (5.4 mm AP, 0 DV, 5.7 LR) was taken as the location of bregma. Craniotomies were targeted accordingly and angles of insertion were set manually. For some of the visual cortex recordings, surface insertion coordinates were targeted based on prior widefield calcium imaging. Widefield imaging used techniques described previously⁴⁴, during presentation of the sparse visual noise receptive field mapping stimulus described above. Responses to visual stimuli near the intended location of the task stimuli were combined and used to identify cortical locations with retinotopically aligned neurons. In some cases, these same imaging sessions were used to target MOp and SSp recordings to the area of large activity observed during forelimb movements that covers both of those areas⁴¹. Finally, MOs recordings were targeted at and around the cortical coordinates identified as disrupting task performance when inactivated, around +2 mm AP, 1 mm ML⁴¹.

Histological probe localization

Recording sites were localized to brain regions by manual inspection of histologically identified recording tracks, in combination with alignment to the Allen Institute Common Coordinate Framework, as follows.

Mice were perfused with 4% paraformaldehyde, the brain was extracted and fixed for 24 h at 4 °C in paraformaldehyde, then transferred to 30% sucrose in PBS at 4 °C. The brain was mounted on a microtome in dry ice and sectioned at 60-μm slice thickness. Sections were washed in PBS, mounted on glass adhesion slides, and stained with DAPI (Vector Laboratories, H-1500). Images were taken at 4× magnification for each section using a Zeiss AxioScan, in three colours: blue for DAPI, green for GCaMP (when present), and red for Dil.

An individual Dil track was typically visible across multiple slices, and recording locations along the track were manually identified by comparing structural aspects of the histological slice with features in the atlas. In most cases, this identification was aided by reconstruction of the track in Allen CCF coordinates. To achieve this, we used the following procedure.

First, we manually identified the 3D locations within the Allen Common Coordinate Framework of each observed Dil spot on each slice (15.1 ± 6.9 such spots per probe, Extended Data Fig. 2a). Code for doing this is provided open-source (<https://github.com/cortex-lab/allenCCF>; in particular, ‘sharp-track’⁴⁵). There was no ambiguity about which dye stain corresponded to which penetration, as penetrations were not repeated within a subject.

After identifying 3D points along a probe penetration, we fit a line to those points, which represents an estimate of the probe trajectory based on all Dil spots (Extended Data Fig. 2b). We quantified the lateral localization error by the median distance of the Dil spots from this common trajectory (39.3 μm) as a quantitative estimate of the error of step no. 1. Though this error is already small, the fitting of a line to many points presumably further reduces the error, resulting in a reliable estimate of the probe’s vector through the brain. Moreover, the predicted set of brain regions that this vector passes through can be directly verified by histological inspection (Extended Data Fig. 3). On completion of this step, the method therefore provides an estimate of the probe trajectory through the 3D atlas, but does not yet provide a mapping from each recording site to a location along this trajectory.

Next, we found the longitudinal mapping from recording sites to the probe trajectory. While the tip location provides one number that can identify the offset, it cannot estimate scaling that may vary owing to shrinkage. We therefore adopted an alternative approach that uses multiple electrophysiological landmarks to estimate scaling and depth on a brain-by-brain basis. These landmarks consist of thin structures such as the CA1 pyramidal layer, or white matter boundaries, which are

Article

unambiguously identifiable in spike rasters. Behavioural task correlates and visual receptive fields were not considered during this alignment procedure. We found the offset and scaling of a linear relationship between recording site number and distance along the probe trajectory by linear regression of these multiple landmarks. We show an example of this process (Extended Data Fig. 2c–e), in which six landmarks are used to find the appropriate depth and scaling (1.01 in this case, and across all recordings 1.08 ± 0.02 (mean \pm s.d.)). We found the shrinkage factor of our fixed tissue relative to the atlas to be 8% on average, but estimated it on a brain-by-brain basis when aligning electrode tracks.

Finally, we quantitatively estimated the accuracy of this longitudinal alignment procedure by taking advantage of the fact there were more points constraining the alignment than there are parameters. To estimate alignment errors, we used a cross-validation approach: we fit the longitudinal mapping using all landmarks except one, and estimated how far the predicted location of this held-out landmark is from its true location (Extended Data Fig. 2f). We found that the median absolute deviation of these errors was 88.7 μm , comparable to the width of the probe shank itself (70 μm), that is, near the resolution limits of the method.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The behavioural and neural datasets generated and analysed in this study are available as downloadable files at <https://figshare.com/articles/steinmetz/9598406> and via the Open Neurophysiology Environment interface at <https://figshare.com/articles/steinmetz/9974357>.

Code availability

The code used to analyse the data are available at <https://github.com/nsteinme/steinmetz-et-al-2019>.

44. Steinmetz, N. A. et al. Aberrant cortical activity in multiple GCaMP6-expressing transgenic mouse lines. *eNeuro* **4**, 0207-17.2017 (2017).
45. Guo, Z. V. et al. Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81**, 179–194 (2014).
46. Bhagat, J. et al. Rigbox: an open-source toolbox for probing neurons and behavior. Preprint at <https://www.biorxiv.org/content/10.1101/672204v3> (2019).
47. Ludwig, K. A. et al. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *J. Neurophysiol.* **101**, 1679–1689 (2009).
48. Park, I. M., Meister, M. L. R., Huk, A. C. & Pillow, J. W. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neurosci.* **17**, 1395–1403 (2014).
49. Ashe, J. & Georgopoulos, A. P. Movement parameters and neural activity in motor cortex and area 5. *Cereb. Cortex* **4**, 590–600 (1994).
50. Paninski, L., Shoham, S., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. Superlinear population encoding of dynamic hand trajectory in primary motor cortex. *J. Neurosci.* **24**, 8551–8561 (2004).
51. Izenman, A. J. Reduced-rank regression for the multivariate linear model. *J. Multivariate Anal.* **5**, 248–264 (1975).
52. Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
53. McGinley, M. J., David, S. V. & McCormick, D. A. Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* **87**, 179–192 (2015).
54. Treves, A. & Rolls, E. T. What determines the capacity of autoassociative memories in the brain? *Netw. Comput. Neural Syst.* **2**, 371–397 (1991).
55. Shamash, P., Carandini, M., Harris, K. D. & Steinmetz, N. A. A tool for analyzing electrode tracks from slice histology. Preprint at *bioRxiv* <https://doi.org/10.1101/447995> (2018).

Acknowledgements We thank A. J. Peters, M. Pachitariu and P. Shamash for software tools; H. Forrest for help with data preprocessing; C. Reddy, M. Wells and L. Funnell for help with mouse husbandry, training and histology; R. Raghupathy for help with histology; C. P. Burgess for help with experimental apparatus; T. Harris and B. Karsh for support with Neuropixels recordings; M. Häusser, A. J. Peters and S. Schroeder for feedback on the manuscript. This project was funded by the European Union's Marie Skłodowska-Curie program (656528), the Human Frontier Sciences Program (LT001071/2015-L), the Wellcome Trust (205093, 102264), the European Research Council (694401), the Gatsby Foundation (GAT3531) and the Simons Foundation (325512). M.C. holds the GlaxoSmithKline / Fight for Sight Chair in Visual Neuroscience.

Author contributions N.A.S., M.C. and K.D.H. conceived and designed the study. N.A.S. collected data. N.A.S., P.Z.-H. and K.D.H. analysed data. N.A.S., M.C. and K.D.H. wrote the manuscript.

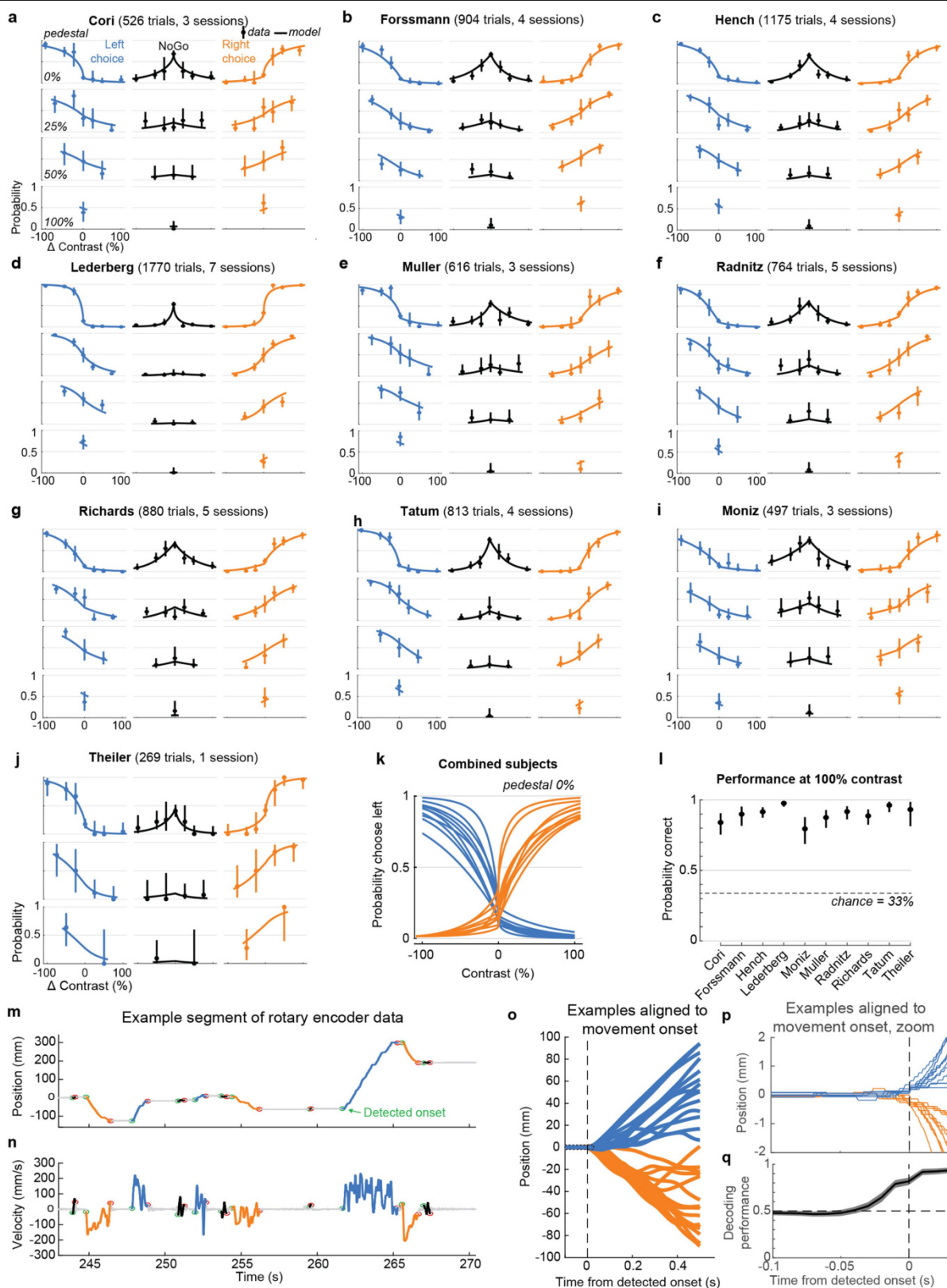
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1787-x>.

Correspondence and requests for materials should be addressed to N.A.S.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

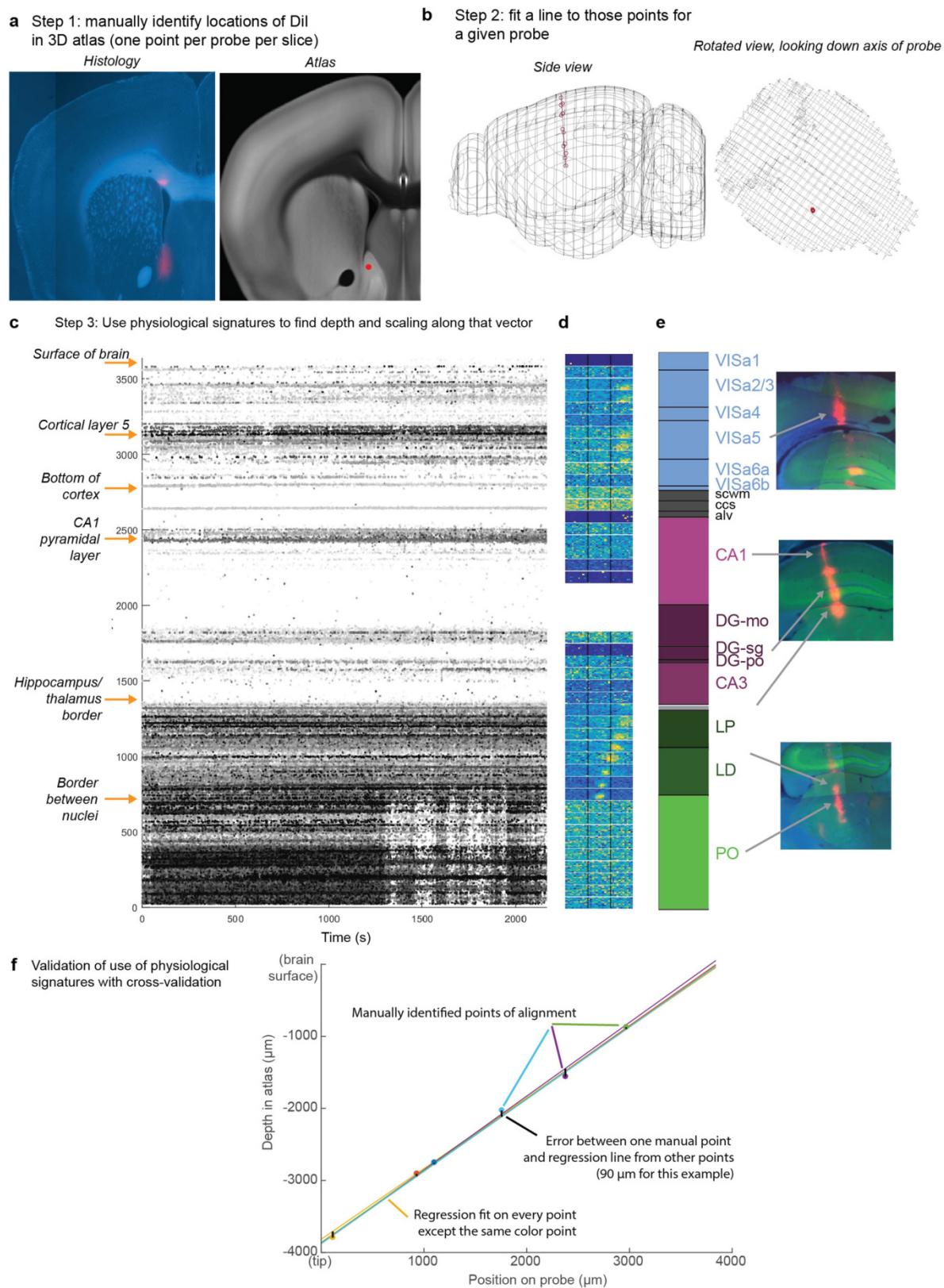


Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | Behavioural performance as psychometric curves for each subject, and analysis of wheel movements. **a**, Psychometric curves for mouse Cori, showing the probability of choosing left (blue), right (orange) or NoGo (black) as a function of stimulus contrasts on the left and right screens. Each row corresponds to a pedestal contrast (the minimum contrast on the left and right screens). The horizontal axis encodes the relative contrast from the pedestal value, positive numbers indicating higher contrast on the right screen, and negative numbers for higher contrast on the left screen (for example, at pedestal = 50%, a Δ Contrast of +50% corresponds to trials with 50% contrast on the left screen and 100% contrast on the right screen). Dots and vertical lines indicate the empirical fraction of choices made and 95% binomial confidence intervals for the fraction estimate, pooling data over sessions. Curves indicate the fit of a multinomial logistic model: $\ln \frac{p(\text{Left})}{p(\text{NoGo})} = b_L + s_L c_L^n$; $\ln \frac{p(\text{Right})}{p(\text{NoGo})} = b_R + s_R c_R^n$, in which c_L and c_R are the contrast on the left and right, and parameters b_L , s_L , n , b_R and s_R are fit by maximum likelihood estimation to the data for each subject³. **b–j**, As in **a**, for the remaining subjects. **k**, The model fit for all subjects overlaid, for left choices (blue) and right choices (orange), in both cases for pedestal = 0%. **l**, Summary of performance on high-contrast

trials. Dots reflect the session-pooled proportion correct of each mouse for trials with 100% versus 0% contrast, with 95% binomial confidence interval. **m**, Example segment of wheel position data showing wheel movements detected as left turns (blue), right turns (orange) or incidental movements (black). Detected onsets (green circles) and offsets (red circles) marked for each movement. y-axis scale: distance moved at the circumference of the wheel (that is, $2\pi R\theta$, in which R is wheel radius and θ is its angular position). **n**, Wheel velocity trace for the same segment of data as in **a**. **o**, Example wheel turns aligned to the detected onset time. The dashed box indicates the region expanded in **p**. **p**, Example wheel turns aligned to detected onset time, zoomed to show the moment of takeoff, illustrating that the wheel had moved by less than 0.5 mm by onset. The step-like appearance of the trace reflects the resolution of the rotary encoder (each step unit is 0.135 mm at the surface of the wheel). **q**, Decoding the eventual direction of the wheel movement using the instantaneous velocity at different times relative to detected movement onset reveals that the direction only starts to be decodable around 20 ms before detected onset, and is not reliably (>80%) decoded until the time of onset. Error bars represent s.d. across sessions ($n = 39$).



Extended Data Fig. 2 | See next page for caption.

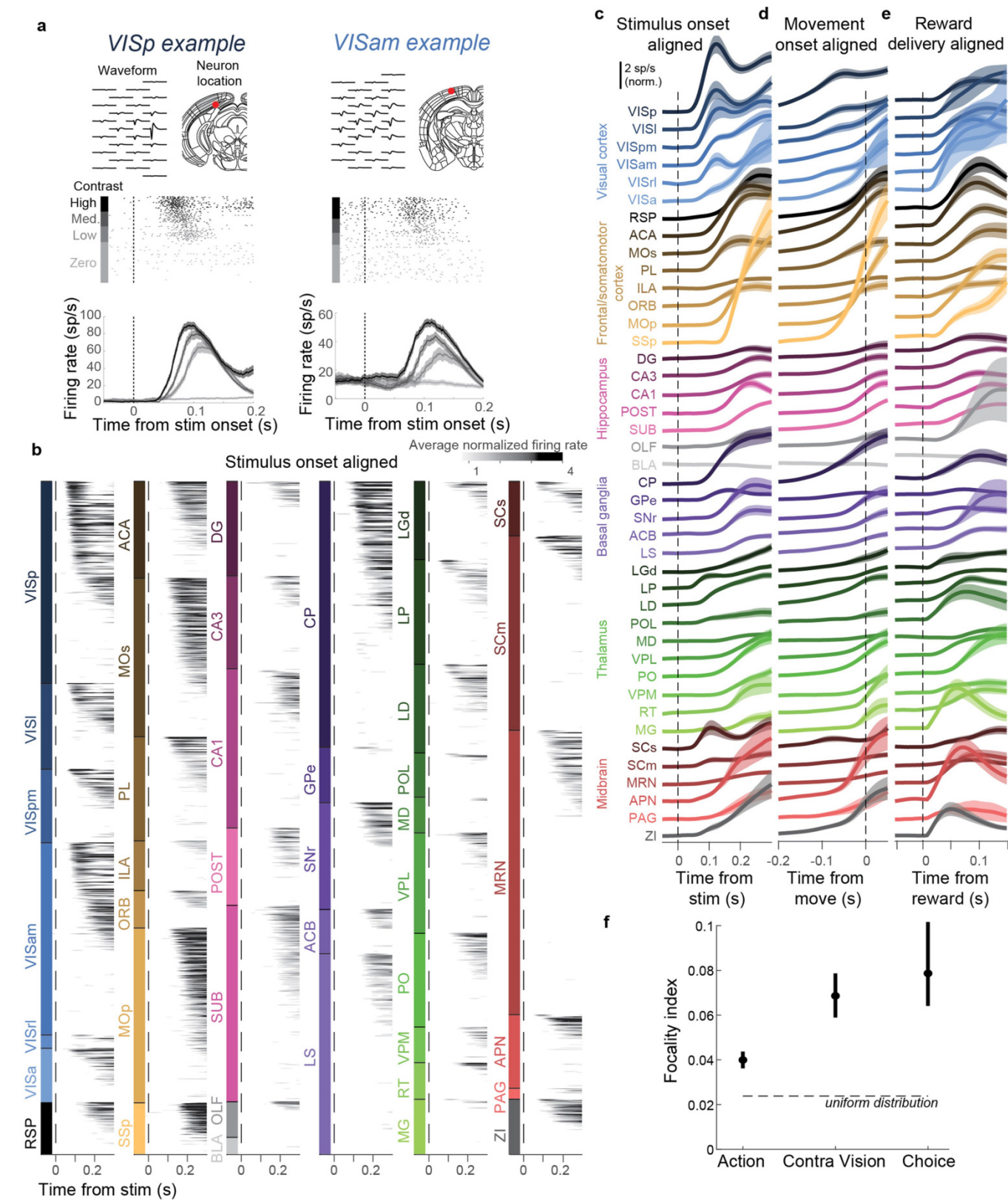
Extended Data Fig. 2 | Method for histological alignment. **a**, Before insertion, probes were coated with Dil. The brain was sliced and imaged, and locations of each probe's Dil spots were manually identified on the Allen CCF atlas (15.1 ± 6.9 spots per probe). When multiple penetrations were performed in a single brain, their tracks were sufficiently far apart to avoid confusion. **b**, A vector is fit to the probe track using total least squares linear regression. The median distance of individual points from this vector is $39.3 \mu\text{m}$, providing an estimate of lateral displacement error. **c**, To fit the longitudinal mapping from recording sites to brain locations, we used landmarks that were easily detectable by their electrophysiological signatures (arrows, left), linearly interpolating the location of sites between these landmarks. **d**, Visual receptive fields served as a post hoc check on correct alignment, but were not used to estimate track location. Each horizontally elongated plot with two vertical black lines

indicates the responsiveness of all spikes recorded in an $80\text{-}\mu\text{m}$ -depth bin to flashed white squares at varying locations on the three screens (see Methods, Receptive field mapping). Colour map indicates spike rate, independently scaled for each map. **e**, Areas assigned for each recording site. Right, example Dil traces in slices corresponding to these locations. **f**, Example of cross-validation procedure to assess error in longitudinal alignment. For each point, the longitudinal mapping was recomputed excluding this point, and the distance from this point to the mapping fit to other points provides an estimate of longitudinal alignment error. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (v.3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).



Extended Data Fig. 3 | Examples of Dil tracks showing recording sites from the depicted sub-surface brain regions in aligned histology. Visual inspection of the Dil tracks confirms that the probe indeed passed through that region at some point along the recording span. Thick white lines outline the region given in the plot title; grey lines outline other regions; and white

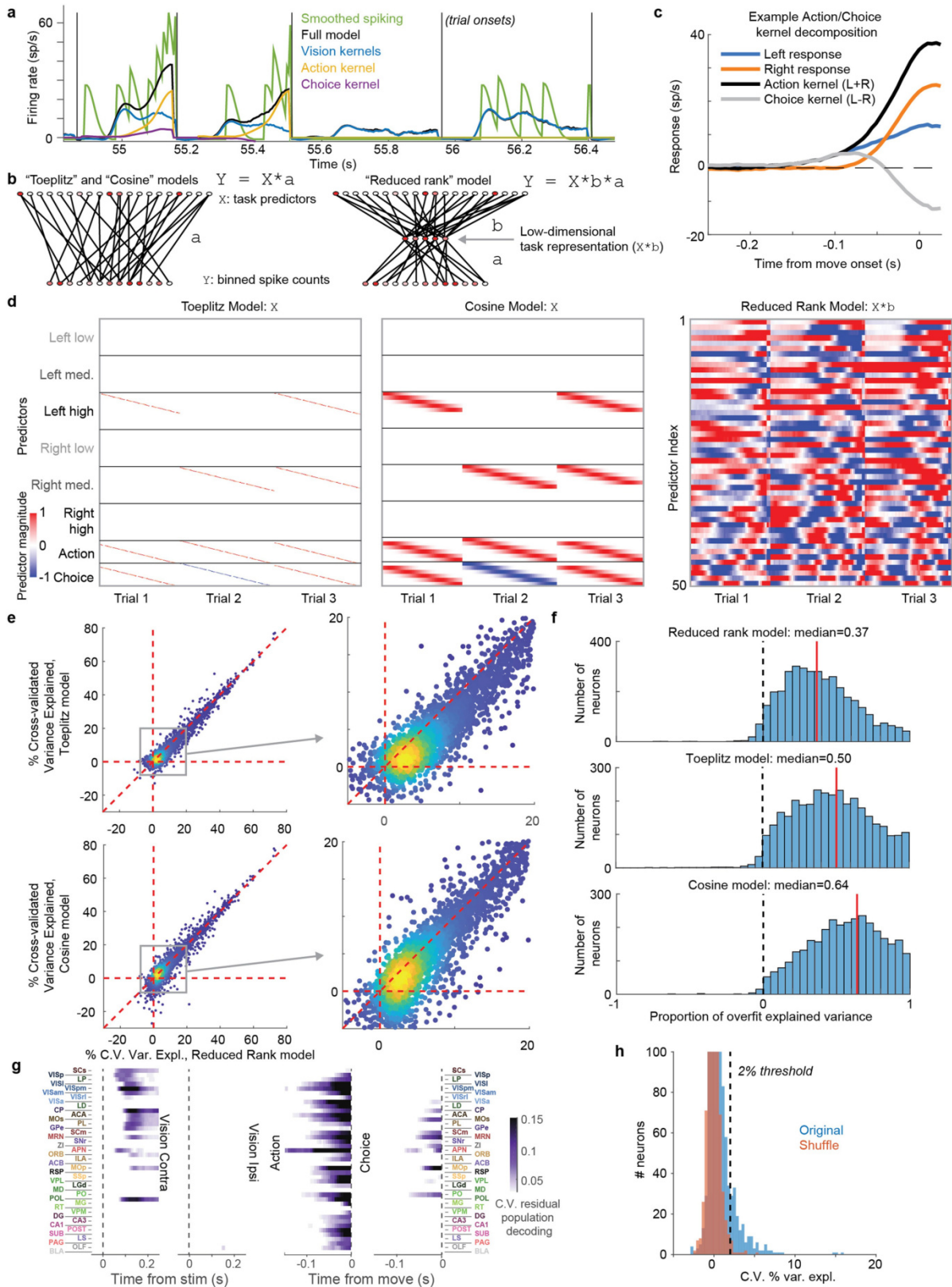
arrows point to places where probe tracks were found within the given region. Blue: DAPI; green: GCaMP; red: Dil. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (v.3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Global neuronal activity and distribution. **a**, Activity of example neurons in VISp and VISam, showing the neuron's waveform and anatomical location (top), rasters sorted by contralateral contrast (middle), and trial-averaged firing rates (smoothed with 30 ms causal half-Gaussian) for each of the four contralateral contrasts (bottom). Shaded regions show the s.e.m. across trials. **b**, Colour map showing trial-averaged firing rates of all highly activated neurons ($P < 10^{-4}$ compared with pre-trial activity), vertically sorted by firing latency. Latency sorting was cross-validated: latencies for each neuron were determined from odd-numbered trials, and activity from even-numbered trials is depicted in the plot. The grey scale represents average normalized firing rate across even-numbered trials with contralateral visual stimuli and contralateral choice. **c–e**, Curves showing mean firing rate across responsive neurons in each area, aligned to visual stimulus onset (**c**), movement onset (**d**) or reward onset (**e**). Trials were included in **e** when a reward

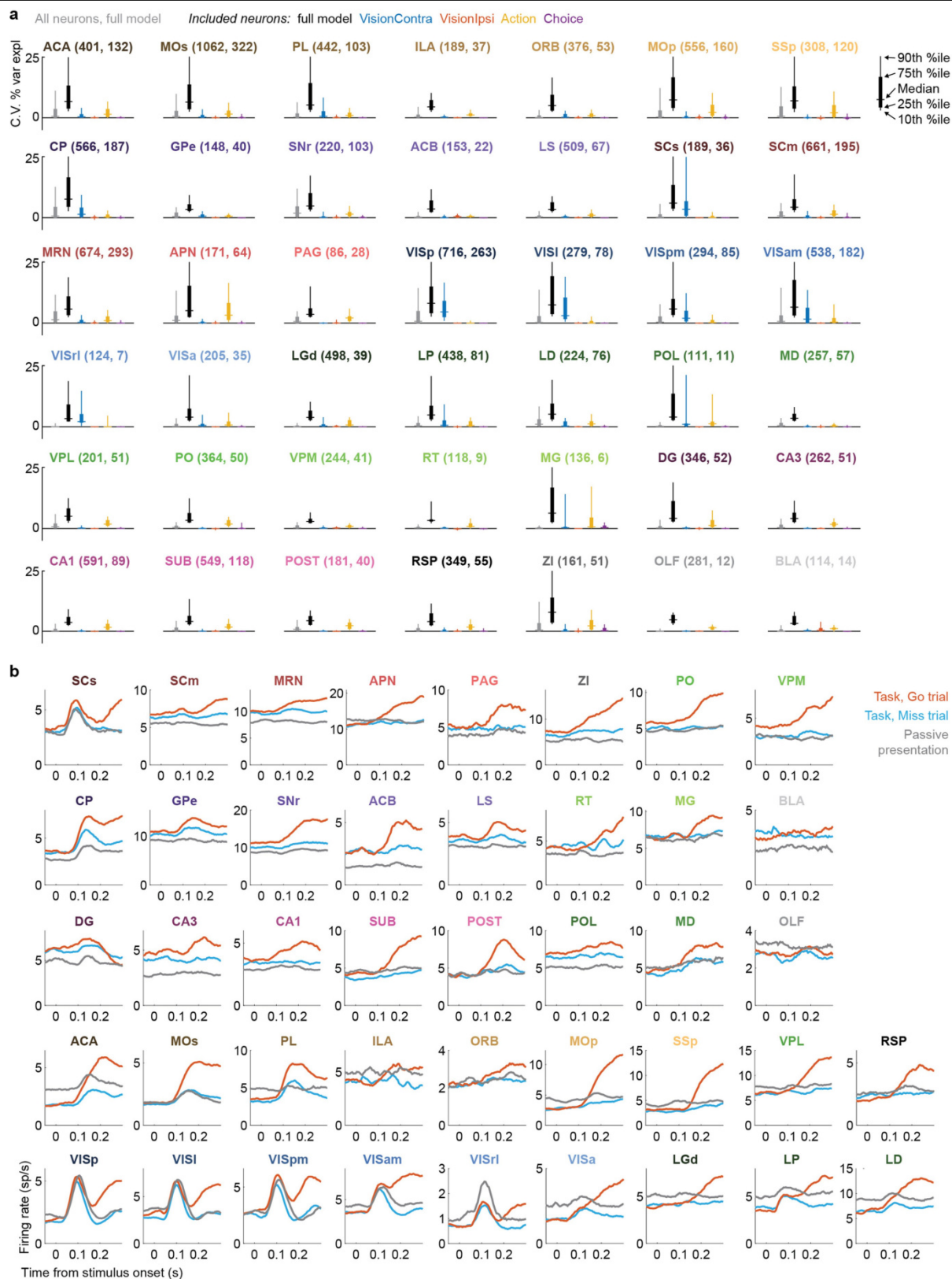
was earned for keeping the wheel still following zero-contrast stimuli. Shaded regions show the s.e.m. across neurons. **f**, The focality index, defined as $\sum (p_a^2) / \sum (p_a)^2$, in which p_a is the proportion of neurons in area a selective for the kernel in question, measures how widely versus focally distributed a representation is, with a floor of 0.0238 for a uniform distribution (across 42 brain regions) and a maximum of 1.0 if all selective neurons were found in a single brain region. This focality index was 0.079 for Choice, 0.069 for Vision kernels and 0.040 for Action kernels; the differences between Choice and Action, as well as contralateral Vision and Action, were statistically significant ($P < 0.05$; bias-corrected bootstrap). Dots represent the true value and error bars represent bias-corrected bootstrap-estimated 95% confidence intervals. Brain diagrams were derived from the Allen Mouse Brain Common Coordinate Framework (v.3 (2017); downloaded from http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/).



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Comparison of the reduced-rank kernel regression method to other methods for spike train prediction. **a**, Example fit of spiking data for an individual neuron with the kernel model. Green trace shows spike data smoothed with a causal filter, black shows the model's prediction, and other coloured traces show the components of the prediction from each kernel. Data between trials is omitted from the fitting and from this plot. **b**, Cartoon of the three methods evaluated. In the Toeplitz and Cosine models, a predictor matrix \mathbf{X} of size $N_{\text{timepoints}} \times N_{\text{predictors}}$ is constructed from task events (illustrated, transposed, in **d**). A linear fit from predictors X to spike counts Y is estimated using elastic net regularization. In the reduced-rank regression method, the predictor matrix \mathbf{X} is the same as the Toeplitz model, but predicts Y after passing through a low-rank bottleneck ($\mathbf{X} \cdot \mathbf{b}$), which is optimized using reduced-rank regression. **c**, Relationship between Action and Choice kernels, which are added or subtracted together to give the shapes for left and right choice trials. This allows separation of neurons with choice from action correlates, while still allowing for arbitrary-shaped responses on left and right trials. **d**, Structure of predictor matrices (shown transposed). The Toeplitz predictor has rows for each variable and time offset, which take non-zero values for time points (columns) corresponding to the appropriate time offset from the given event. The cosine model has similar structure but with rows replaced by smooth raised cosine functions, allowing a smaller number of basis

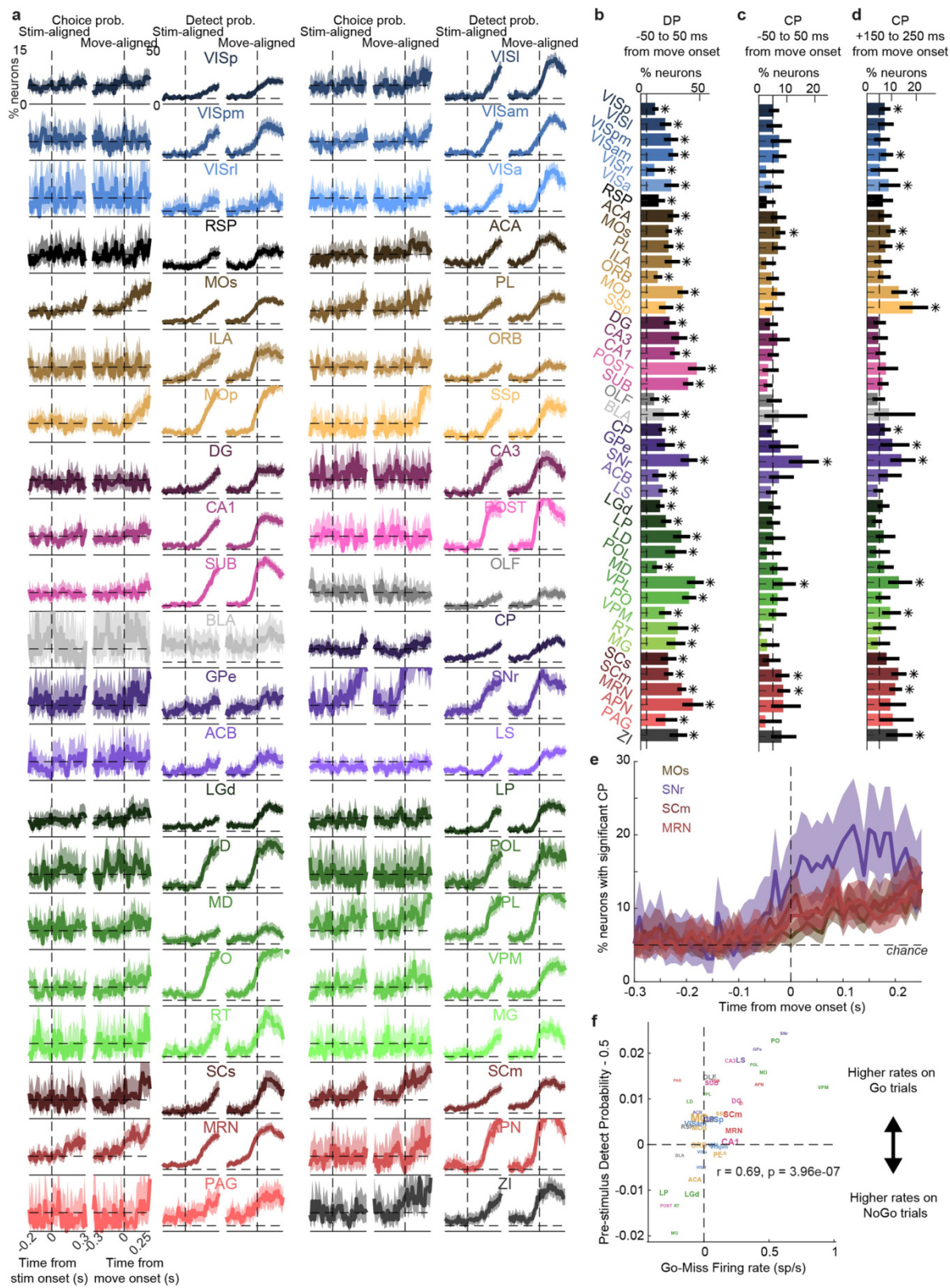
functions. The reduced-rank regression model has learned a small number of dense basis functions optimized to predict spike counts. **e**, Density scatterplot of cross-validated variance explained for each neuron under the Toeplitz model against the reduced-rank model (top), and for the cosine model versus the reduced-rank model (bottom). Each point represents one cell, coloured to show density when they overlap. Right, magnified view of the densest region of the plot. These comparisons show that the reduced-rank model consistently outperforms the other two (points lie below the diagonal), and that it overfits fewer neurons (fewer points with CV variance explained < 0). **f**, The proportion of overfit-explained variance, that is, $(CV_{\text{train}} - CV_{\text{test}})/CV_{\text{train}}$, in which CV_{train} is the train-set variance explained and CV_{test} is the test-set variance explained. Smaller values for the reduced-rank model show that it overfits less. **g**, Left, population decoding of contralateral visual stimulus contrast from residual population activity in each area after subtracting the prediction of a model including all other kernels. The other panels depict the same analysis for decoding of ipsilateral visual stimulus contrast, action and direction of choice. **h**, Distributions of the cross-validated proportion variance explained for each neuron when shuffling left and right trial choice labels (orange) together with the distribution for the original data (blue). After shuffling, a small number (14, 0.33%) of neurons are false positives by this threshold. The dashed line represents the 2% CV variance explained threshold used. The y axis is clipped.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Summary of variance explained by the kernel model and population average responses on Go, Miss, and Passive trials. **a**, The unique contribution of each predictor variable as assessed by nested prediction. Each panel depicts the distribution of variance explained across neurons of a single brain region, using various reduced-rank kernel regression models, (compare with Figs. 3c, e, 4b). Each bar shows the 10th, 25th, 50th, 75th and 90th percentiles of the distribution for a single prediction model, colour-coded by model identity. The numbers in the subplot title indicate the number of all neurons analysed with the full model (that is, the distribution shown with the grey bar), and the number of neurons included for nested model analysis (that is, cells with $\geq 2\%$ variance explained with the full model). The black bar shows distribution of variance explained by the full model in this subset; coloured bars show the unique contribution of each predictor. Note that the

unique contributions need not sum to the variance of the full model, as predictor variables are correlated. Variance explained by the Action kernel (yellow) is essentially global, whereas contralateral Vision variance explained is distinctly restricted, and Choice is rare enough to be difficult to see in these plots. **b**, Population average firing rates across neurons for each brain region in Go, Miss, and passive trials, selected to have matched contralateral visual stimulus contrasts. The patterns characteristic of engagement can be seen in pre-stimulus activity (that is, before time 0): the pre-stimulus firing rate of midbrain, basal ganglia, and hippocampal regions is in the order Go > Miss > passive; whereas pre-stimulus activity in neocortical areas instead is arranged as passive > Miss and Go. Thalamic regions can exhibit either pattern; notably, visual thalamic regions (LGd, LP and LD) follow the pattern of neocortical areas.

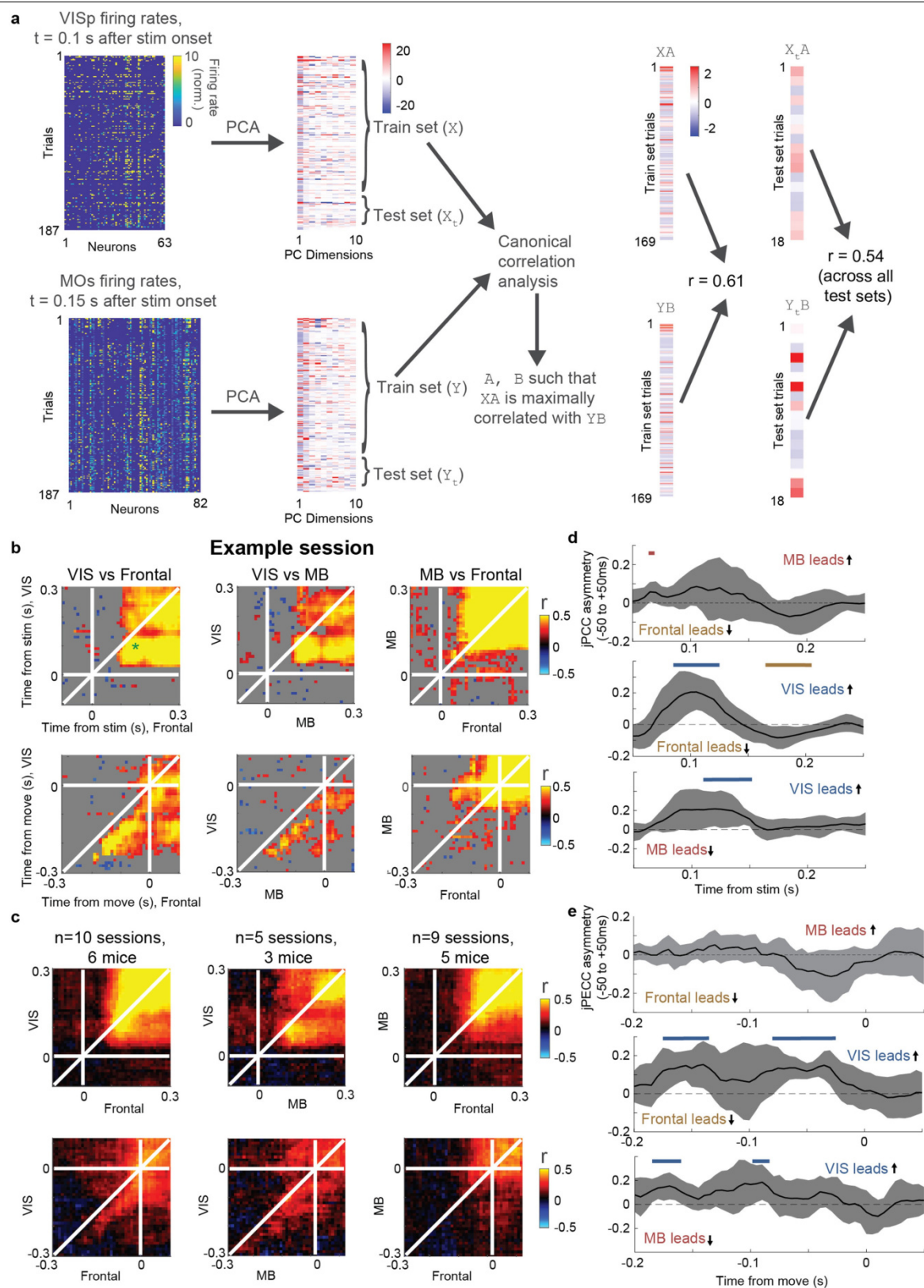


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Choice probability and detect probability analysis. a,

The percentage of neurons with significant ccCP (that is, neurons whose rate differed significantly between left and right choices in response to the same stimulus; left two columns) and ccDP (that is, neurons whose rate differed significantly between Go and NoGo trials in response to the same stimulus; right two columns) as a function of time aligned to visual stimulus onset (left) and movement onset (right). The horizontal dashed line represents the value expected by chance given the statistical threshold $\alpha = 0.05$. **b**, Percentage of neurons in each area with significant detect probability between -50 and $+50$ ms from movement onset, replicating the finding from Fig. 2d, e, 3e that non-selective action signals are distributed widely. Asterisks indicate brain regions for which 95% confidence intervals for the percentage of significant neurons (black error bars) did not include the chance value (5%, horizontal dashed line). **c**, As in **b**, for choice probability in the same window. The number of trials usable in this analysis is limited, meaning that some sessions ($n = 6$ of 39) had to be excluded; nevertheless, this analysis broadly replicates the finding from Fig. 4b that around the time of movement onset, choice-selective neurons are restricted to frontal cortex, basal ganglia, midbrain and certain thalamic nuclei. **d**, As in **a**, for choice probability in a window $+150$ to 250 ms

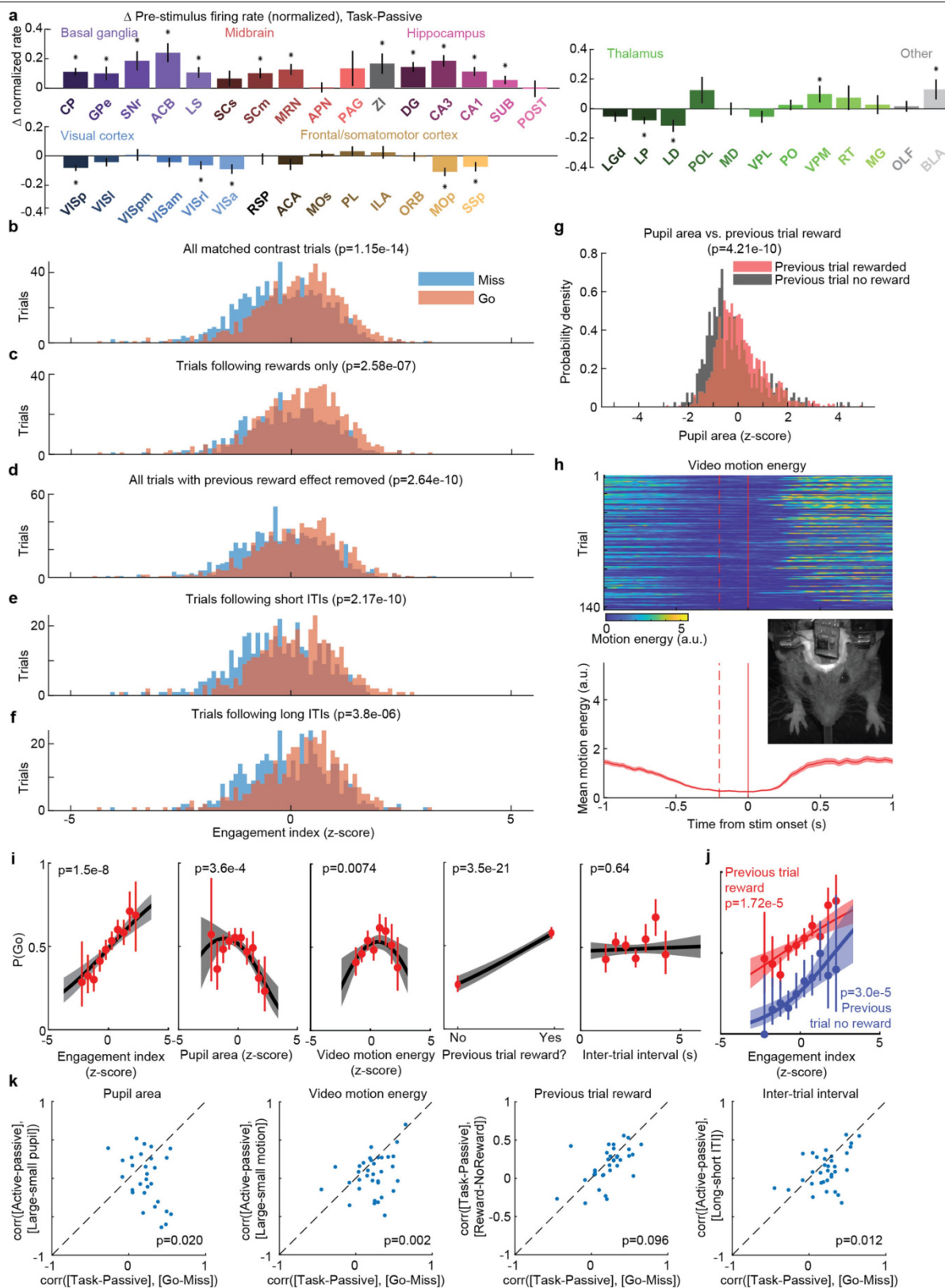
after movement onset, showing that by this time choice-related signals are distributed more widely, including visual and parietal cortex. These signals are too late to have participated in generating the choice but could reflect either corollary discharge or sensory reafference. However, they cannot reflect movement of the visual stimulus on the screen, as it is fixed during this time period. **e**, The percentage of neurons with significant choice probability, as a function of time relative to movement onset for selected areas (zoom and overlay of certain traces from **a**), replicating that choice related activity is first seen in the final 50 – 100 ms relative to movement onset and with similar timing across multiple areas. Note that six sessions were excluded from ccCP analysis for having too few trials; these six sessions included 20.8% of the MOs neurons determined to have choice-selective responses with kernel regression. **f**, The pre-stimulus detect probability (after subtracting 0.5, so that positive values indicate higher rates on Go trials, and negative values the reverse) versus the mean Go – Miss firing rate difference for each area (used in Fig. 5d), demonstrating that these two quantities identify essentially the same factor. The pre-stimulus detect probability was correlated with the engagement index (that is, task – passive difference) similarly to the Go – Miss difference (with $r = 0.48$, $P = 0.001$; not shown).



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | jPECC analysis for determining whether correlations occur with a temporal offset between a pair of regions. **a**, Canonical correlation analysis is applied to firing rates at every pair of time points relative to a behavioural event (illustration shows 0.1 s after stimulus onset in VISp and 0.15 s after in MOs). Canonical correlation analysis is applied to the pair of matrices containing each cell's firing rate at selected times on each training set trial (90% of the total) to find dimensions in each population maximally correlated with each other. (For regularization purposes, this is applied after dimensionality reduction using principal component analysis). The strength of population correlation is summarized by the correlation of test set activity projected onto the first canonical dimension. **b**, Results on an example session, showing relationships between visual cortex, midbrain and frontal cortex

relative to stimulus onset (top) and movement onset (bottom). Visual cortical activity leads frontal and midbrain activity, as can be seen from the below-diagonal bias in correlations. However, no lead/lag relationship is seen between midbrain and frontal cortex. Grey, $P > 0.05$. **c**, Average across all recording sessions that contained each pair of areas, showing similar relationships to the example in each case. **d, e**, Summary of lead-lag interactions, obtained by subtracting the averages of the jPECC coefficients over inter-area time ranges of -50 to 0 and 0 to 50 ms, as a function of time relative to stimulus onset (**d**) or movement onset (**e**). Grey region, $2 \times \text{s.e.m.}$ across experiments. Visual cortex reliably leads frontal cortex and midbrain at around 100 ms after the stimulus; and over a range -200 to -50 ms relative to movement.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Statistical analysis of engagement index and influence of alertness-, reward- and history-related variables on pre-stimulus firing rates.

a, A nested ANOVA with factors of session and subject was used to assess statistical significance of pre-stimulus task-passive firing rate differences (here normalized, unlike Fig. 5c, d and Extended Data Fig. 6b, to meet statistical assumptions) in each brain region (see Methods). All non-neocortical regions that showed a significant difference between engaged task and passive states had higher mean pre-stimulus firing rates in task context, except for visual thalamus. All neocortical regions that showed a significant difference between task and passive contexts had lower mean pre-stimulus firing rates in the task context. **b**, An engagement index was computed for each task trial by projecting pre-stimulus population activity onto the vector of differences between pre-stimulus activity in task and passive contexts. Task trials with lower engagement index therefore showed a pattern of pre-stimulus population activity more similar to the passive context. Histogram shows the distributions of this index over contrast-matched Miss and Go trials; *P* value computed by *t*-test. **c–f**, Same plot after restricting to contrast-matched trials following rewards (**c**), after removing the reward effect by partial regression (that is, by subtracting the mean within trials of each previous reward condition) (**d**), after restricting to contrast-matched trials following short inter-trial intervals (**e**) or after restricting to contrast-matched trials following long inter-trial intervals (**f**). The effect persists in each case. **g**, Histogram of pupil areas in the pre-stimulus period after previous trials that were rewarded or non-rewarded, showing the expected effect of reward on arousal as a positive control for the validity of pupil diameter measurements. **h**, To initiate the next trial, subjects must hold the wheel still for 500 ms; video analysis shows that they reduce other movements as well. Top, total video motion energy (mean-square frame difference) as a function of time relative to stimulus onset, on

each trial, for an example recording. Bottom, mean motion energy across these trials overlaid (red-shaded region represents s.e.m. across trials). Inset, example frame from video monitoring the face and forelegs of the mouse. **i**, Results of a logistic GLM predicting the probability of a Go response on the subsequent trial from each of the given variables; plot format as in Fig. 5e. The null hypothesis that engagement index had no additional effect on Go probability over all other variables was rejected using a deviance test ($P = 1.5 \times 10^{-8}$). Each panel's curve shows effect of one individual variable on Go probability. Red points, mean Go probability averaged over a bin; red error bars, 95% confidence interval; black line, fit of GLM, setting all other variables to their mean; grey shading, 95% confidence interval. **j**, Average Go probability and GLM fit as a function of engagement index (*x* axis), and previous trial reward (colour). Correlation of *P*(Go) from engagement index persists despite the additional effect of previous trial's reward. **k**, To additionally test whether engagement index more specifically relates to *P*(Go) than any other variable, we asked whether the engagement vector (that is, the mean difference in pre-stimulus population activity between task and passive contexts) matches the population vector encoding differences before Go and Miss trials, better than it does differences in other behavioural variables. To do so, we computed the Pearson correlation between the engagement vector and the Go–Miss vector for each recording. This correlation coefficient can be interpreted as the cosine of the angle between the two vectors in N_{neurons} dimensional space. Each panel shows a scatter plot, with one dot per recording, comparing this correlation against the engagement vector's correlation with vectors defined for each other behavioural variable. In each case we find that the correlation is greater for the Go–Miss vector (not reaching significance at $\alpha = 0.05$ level, however, for the comparison with the previous reward vector, $P = 0.096$).

Extended Data Table 1 | Brain regions recorded

Abbreviation	Full name	N recordings	N mice	N responsive neurons	N total neurons	% responsive
ACA	Anterior cingulate area	11	8	395	608	65.0
ACB	ACB Nucleus accumbens	3	3	152	255	59.6
APN	APN Anterior pretectal nucleus	2	2	177	235	75.3
BLA	Basolateral amygdalar nucleus	2	2	107	261	41.0
CA1	Field CA1	21	10	494	1129	43.8
CA3	Field CA3	10	6	243	438	55.5
CP	Caudoputamen	5	5	524	914	57.3
DG	Dentate gyrus	16	10	336	711	47.3
GPe	Globus pallidus external segment	3	3	146	274	53.3
ILA	Infralimbic area	3	3	192	338	56.8
LD	Lateral dorsal nucleus of the thalamus	6	5	209	308	67.9
LGd	Dorsal part of the lateral geniculate complex	10	5	397	811	49.0
LP	Lateral posterior nucleus of the thalamus	11	8	393	732	53.7
LS	Lateral septal nucleus	7	5	508	844	60.2
MD	Mediodorsal nucleus of the thalamus	3	3	244	381	64.0
MG	Medial geniculate complex of the thalamus	2	2	180	276	65.2
MOp	Primary motor area	3	3	553	682	81.1
MOs	Secondary motor area	19	9	993	1534	64.7
MRN	Midbrain reticular nucleus	11	7	722	857	84.2
OLF	Olfactory areas	9	5	210	684	30.7
ORB	Orbital area	6	5	281	770	36.5
PAG	Periaqueductal gray	3	3	72	130	55.4
PL	Prelimbic area	10	7	438	728	60.2
PO	Posterior complex of the thalamus	5	4	342	620	55.2
POL	Posterior limiting nucleus of the thalamus	3	3	132	190	69.5
POST	Postsubiculum	4	3	163	272	59.9
RSP	Retrosplenial area	9	5	308	598	51.5
RT	Reticular nucleus of the thalamus	2	2	111	160	69.4
SCm	Superior colliculus motor related	11	6	738	997	74.0
SCs	Superior colliculus sensory related	10	6	192	317	60.6
SNr	Substantia nigra reticular part	4	3	201	274	73.4
SSp	Primary somatosensory area	5	4	296	461	64.2
SUB	Subiculum	9	7	494	669	73.8
VISa	Anterior visual area	5	4	207	285	72.6
VISam	Anteromedial visual area	11	8	501	805	62.2
VISl	Lateral visual area	3	2	248	403	61.5
VISp	Primary visual area	12	8	649	923	70.3
VISpm	Posteromedial visual area	4	4	230	516	44.6
VISrl	Rostrolateral visual area	2	2	85	236	36.0
VPL	Ventral posterolateral nucleus of the thalamus	4	3	202	255	79.2
VPM	Ventral posteromedial nucleus of the thalamus	4	2	235	357	65.8
ZI	Zona incerta	4	2	167	220	75.9

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Matlab; <https://github.com/cortex-lab/Rigbox>; <http://billkarsh.github.io/SpikeGLX/>

Data analysis

Matlab; <https://github.com/cortex-lab/KiloSort>; <https://github.com/cortex-lab/spikes>; <https://github.com/kwikteam/phy>; <https://github.com/nsteinme/steinmetz-et-al-2019>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

DOI: 10.6084/m9.figshare.9598406

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences
- ☐ Behavioural & social sciences
- ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n=10 mice were used because this was a sufficient number to record n=29134 neurons in n=42 brain regions, numbers which are greater than any previously published work. No statistical calculations were used to determine exactly how many neurons to record.
Data exclusions	Behavioral trials were excluded when there were three or more miss trials in a row indicating disengagement; this criterion was pre-established. Neurons were excluded when the null hypothesis that their firing rate was un-modulated during the task could not be rejected according to a series of four tests for different types of modulation; this criterion was not determined in advance but was developed from data inspection.
Replication	Recordings were undertaken in n=10 mice and a population of neurons recorded in each brain region in at least two of the mice.
Randomization	Trial types during the behavioral task were interleaved randomly.
Blinding	n/a

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Experiments were performed on male and female mice, between 11 and 46 weeks of age. Multiple genotypes were employed, including: Ai95;Vglut1-Cre (B6J.Cg-Gt(ROSA)26Sortm95.1(CAG-GCaMP6f)Hze/MwarJ crossed with B6;129S-Slc17a7tm1.1(cre)Hze/J), TetO-G6s;Camk2a-tTa (B6;DBA-Tg(tetO-GCaMP6s)2Niell/J crossed with B6.Cg-Tg(Camk2a-tTA)1Mmay/DboJ), Snap25-G6s (B6.Cg-Snap25tm3.1Hze/J), Vglut1- Cre, and wild-type (C57Bl6/J).
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from a field.
Ethics oversight	Experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986) and under personal and project licenses released by the Home Office following appropriate ethics review.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

The molecular landscape of ETMR at diagnosis and relapse

<https://doi.org/10.1038/s41586-019-1815-x>

Received: 21 December 2018

Accepted: 16 October 2019

Published online: 4 December 2019

Sander Lambo^{1,2,3}, Susanne N. Gröbner^{1,2,3}, Tobias Rausch⁴, Sebastian M. Waszak⁴, Christin Schmidt^{1,2,3}, Aparna Gorthi^{5,6}, July Carolina Romero^{5,6}, Monika Mauermann^{1,2,3}, Sebastian Brabetz^{1,2,3}, Sonja Krausert^{1,2,3}, Ivo Buchhalter⁷, Jan Koster⁸, Danny A. Zwiijnenburg⁸, Martin Sill^{1,2,3}, Jens-Martin Hübner^{1,2,3}, Norman Mack^{1,2,3}, Benjamin Schwalm^{1,2,3}, Marina Ryzhova⁹, Volker Hovestadt¹⁰, Simon Papillon-Cavanagh¹¹, Jennifer A. Chan¹², Pablo Landgraf¹³, Ben Ho¹⁴, Till Milde^{1,3,15,16}, Olaf Witt^{1,3,15,16}, Jonas Ecker^{1,15,16}, Felix Sahm^{1,3,17,18}, David Sumerauer¹⁹, David W. Ellison²⁰, Brent A. Orr²⁰, Anna Darabi²¹, Christine Haberler²², Dominique Figarella-Branger^{23,24}, Pieter Wesseling^{25,26}, Jens Schittenhelm^{27,28}, Marc Remke^{3,29,30}, Michael D. Taylor³⁰, Maria J. Gil-da-Costa³¹, Maria Łastowska³², Wiesława Grajkowska³², Martin Hasselblatt³³, Peter Hauser³⁴, Torsten Pietsch³⁵, Emmanuelle Uro-Coste^{36,37}, Franck Bourdeaut³⁸, Julien Masliah-Planchon^{39,40}, Valérie Rigau^{41,42}, Sanda Alexandrescu⁴³, Stephan Wolf⁴⁴, Xiao-Nan Li^{45,46,47}, Ulrich Schüller^{48,49,50}, Matija Snuderl⁵¹, Matthias A. Karajannis⁵², Felice Giangaspero^{53,54}, Nada Jabado¹¹, Andreas von Deimling^{1,17,18}, David T. W. Jones^{1,3,55}, Jan O. Korbel⁴, Katja von Hoff^{56,57}, Peter Lichter^{3,10}, Annie Huang^{14,58}, Alexander J. R. Bishop^{5,6,59}, Stefan M. Pfister^{1,2,3,16,60}, Andrey Korshunov^{17,18,60} & Marcel Kool^{1,2,3,60*}

Embryonal tumours with multilayered rosettes (ETMRs) are aggressive paediatric embryonal brain tumours with a universally poor prognosis¹. Here we collected 193 primary ETMRs and 23 matched relapse samples to investigate the genomic landscape of this distinct tumour type. We found that patients with tumours in which the proposed driver C19MC^{2–4} was not amplified frequently had germline mutations in *DICER1* or other microRNA-related aberrations such as somatic amplification of *miR-17-92* (also known as *MIR17HG*). Whole-genome sequencing revealed that tumours had an overall low recurrence of single-nucleotide variants (SNVs), but showed prevalent genomic instability caused by widespread occurrence of R-loop structures. We show that R-loop-associated chromosomal instability can be induced by the loss of *DICER1* function. Comparison of primary tumours and matched relapse samples showed a strong conservation of structural variants, but low conservation of SNVs. Moreover, many newly acquired SNVs are associated with a mutational signature related to cisplatin treatment. Finally, we show that targeting R-loops with topoisomerase and PARP inhibitors might be an effective treatment strategy for this deadly disease.

ETMR is a malignant type of brain tumour that occurs almost exclusively in young children¹. The tumours show diverse histological patterns that have been described as ependymoblastoma (EBL), medulloepithelioma (MEPL) or embryonal tumour with abundant neuropil and true rosettes (ETANTR), which together form one distinct biological tumour type termed ETMR^{1,5}. Genetically, ETMRs are characterized by amplification and fusion of a microRNA (miRNA) cluster on chromosome 19 (C19MC) with *TTYH1*, which is present in approximately 90% of the tumours^{2–4}. As current treatment options fail for nearly all patients, a better understanding of ETMR tumorigenesis is needed to develop new treatment strategies. Here we characterized the molecular landscape of a large cohort of patients with primary and relapsed ETMRs to gain more insight into the inter- and intratumoral heterogeneity of ETMRs at diagnosis and relapse, and to identify tumour-driving mechanisms that may lead to more-effective treatment strategies.

Intertumoral heterogeneity

Cluster analysis of ETMR DNA methylation profiles ($n = 193$; Supplementary Table 1) or mRNA data ($n = 28$), confirmed that ETMRs are clearly distinct from other brain tumour types^{6,7} (Fig. 1a, b). ETMRs without C19MC amplification ($n = 23$) tend to cluster together at the edge of the main ETMR cluster but do not really separate, even when clustering only ETMRs (Extended Data Fig. 1a). Additionally, miRNAs are expressed in a distinct pattern in ETMRs, but are similar between ETMRs with ($n = 7$) and without ($n = 3$) C19MC amplification (Extended Data Fig. 2). C19MC miRNAs are also expressed in ETMRs without C19MC amplification, albeit at an approximately tenfold lower level, but not in normal brain or other brain tumours (Fig. 1c). Mature miRNAs that are specifically upregulated in ETMRs included all C19MC miRNAs and members of the *miR-17-92* miRNA cluster, whereas several members of the let-7 family of miRNAs are specifically downregulated in ETMRs (Supplementary Table 2).

A list of affiliations appears at the end of the paper.

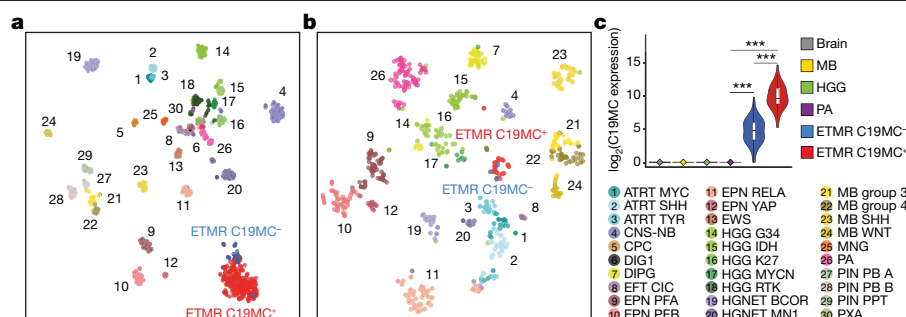


Fig. 1 | ETMRs regardless of C19MC amplification show high molecular similarity. **a, b**, *t*-Stochastic neighbourhood embedding clustering using either DNA methylation (**a**) or mRNA expression (**b**) data. Colours represent different tumour entities classified as described previously⁷. Definitions of the groups are provided in Supplementary Table 10. **c**, Violin plot showing \log_2 -

transformed expression of C19MC miRNAs ($n = 56$) in ETMRs and other tissues. *P* values were calculated using two-sided Mann–Whitney *U*-tests; *** $P < 0.0005$. Box plots show the median \pm interquartile range, whiskers extend to $1.5 \times$ the interquartile range, violin plots depict kernel density estimates and represent the density distribution.

ETMRs occur throughout the brain, without any association with different histological variants, although tumours without C19MC amplification were significantly more often located in infratentorial regions (Fisher's exact test, $P = 8.80 \times 10^{-7}$). None of the other clinical annotations was associated with any molecular subgrouping (Extended Data Fig. 1a, b). Taken together, these data show that ETMRs are a molecularly distinct tumour type, with limited molecular intertumour heterogeneity despite diverse histopathological features, absence of C19MC amplification in a subset of tumours and a variety of anatomical locations.

Intratumour heterogeneity

By comparing ETMR mRNA data against healthy brain tissue and other brain tumours (Supplementary Table 3), we identified several upregulated pathways that are related to development, including RNA processing, and Hippo, WNT, Notch and SHH signalling (Extended Data Fig. 3). Additionally, compared to other tumours, most ETMRs showed upregulation of DNA repair pathways, including base excision repair, Fanconi anaemia and the p53 pathway⁸ (Extended Data Fig. 4a and Supplementary Table 4). However, a subset of tumours (10 out of 28) lacked this upregulation. As no distinct molecular subgroups were detected, we investigated whether this was due to differences in intratumour heterogeneity. CIBERSort analysis was used to compare tumour expression data with single-cell RNA-sequencing data of the developing prefrontal cortex to delineate distinct tumour cell populations^{9,10}. This resulted in the identification of stem-cell-like, more-differentiated oligodendrocyte precursor-like and astrocyte-like tumour cells (Extended Data Fig. 4b). A higher proportion of differentiated progeny corresponded to lower levels of the stem cell markers *LIN28A* and *HMG2A*, lower levels of the DNA-damage checkpoint genes *WEE1* and *CHEK2*, but higher levels of the astrocyte markers *AQP4* and *GFAP* (Extended Data Fig. 4c).

Notably, these observations correlated with the histology of the tumours. Tumours with more-differentiated cells and lower expression of DNA repair genes tended to be more often diagnosed as ETANTRs, while tumours with less-differentiated cells and higher expression of DNA repair genes were more likely to be described as EBL or MEPL (Extended Data Fig. 4d). These findings were validated by methylation profiling and RNA sequencing of microdissected neuropil (more differentiated) and rosettes (less differentiated) of an ETMR diagnosed as ETANTR. Although RNA sequencing confirmed that the expression of stem cell and astrocyte markers differed between the two cell populations, methylation profiles were highly similar and clustered together with other ETMRs (Extended Data Fig. 4e–g and Supplementary Table 3).

Previously, we reported that *LIN28A* is more widely expressed upon relapse and histology shifts towards a more EBL or MEPL phenotype¹.

To investigate whether this is due to the outgrowth of the stem-cell-like population after treatment, we compared the expression of *LIN28A*, *HMG2A*, *WEE1*, *CHEK2*, *AQP4* and *GFAP* between a primary ETANTR and two matched recurrences and found that in both relapse samples the gene expression is shifting to be more stem-cell-like (Extended Data Fig. 4h). Differences in expression of DNA repair genes and histology are therefore likely to be explained by the levels of differentiated progeny, while undifferentiated cells with high levels of DNA repair expression grow out upon relapse.

Recurrent aberrations

We sequenced 82 tumours, including 16 cases without C19MC amplification and 12 recurrences (Methods and Supplementary Tables 5, 6), to see whether there were any recurrent DNA aberrations other than C19MC amplification. Other recurrently mutated genes that were detected in the primary tumours included *DICER1* (8 out of 70), *CTNNB1* (7 out of 70) and *TP53* (5 out of 70), whereas mutations (other than *CTNNB1*) in developmental pathways, which have previously been suggested to have a role in ETMR tumorigenesis¹¹, were detected only sporadically (Fig. 2a).

Notably, all *DICER1* mutations occurred in cases that lacked the C19MC amplification. All eight samples had compound heterozygous mutations, including one somatic mutation (E1705K (three samples), D1709N (two samples), G1809R, E1813G and E1813K) in the hotspot region that affects the RNASE IIIb domain, which is important for miRNA processing¹². The second mutation was found in the germline (validated for seven out of eight samples) and was in all cases deleterious (Fig. 2b). Sequencing of mature miRNAs indeed showed a strongly increased ratio of 3' over 5' mature miRNAs in one of the *DICER1*-mutated samples (ET31) compared to individuals that have no *DICER1* mutations (Fig. 2c).

Two ETMRs (ET68 and ET173, which lacked C19MC amplification and *DICER1* mutations) had amplifications of the *miR-17-92* miRNA cluster on chromosome 13. In ET68 this cluster was fused to a region on chromosome 11 to form a circular chromosome, which was confirmed by mate pair sequencing, copy-number profiles and fluorescence in situ hybridization (Extended Data Fig. 5a, b). Genes on the circular chromosome, including the *miR-17-92* miRNA cluster, showed an approximately fourfold higher expression level compared to other ETMRs (Extended Data Fig. 5c). The amplification of the *miR-17-92* miRNA cluster was also found once in an ETMR (ET85) with C19MC amplification.

Three other samples (ET89, ET160 and ET168, which lacked C19MC amplification and *DICER1* mutations) showed clustered breakpoints that affected the C19MC locus, suggesting that even though C19MC was not amplified, its expression might be driven by structural rearrangements (Extended Data Fig. 5d). In ET160, we found a fusion between C19MC and *MYO9B*, which is also located on chromosome 19. Alternative

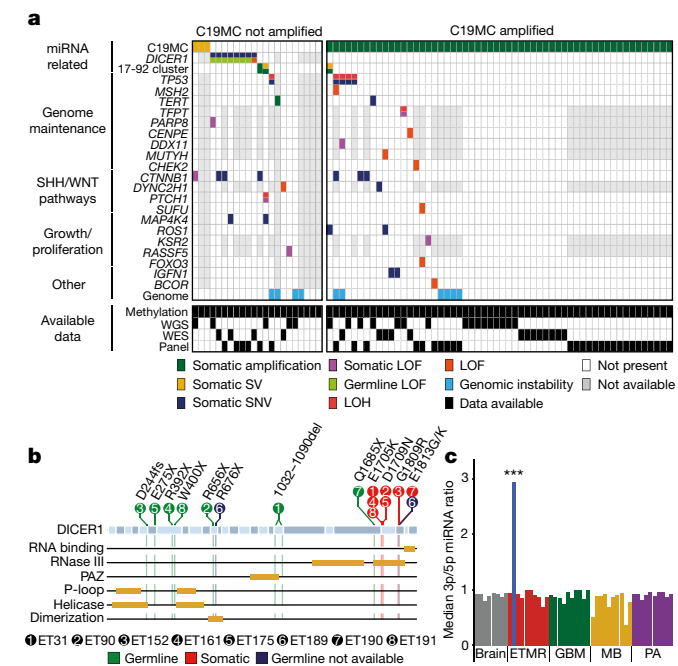


Fig. 2 | ETMRs without C19MC amplification recurrently have miRNA related aberrations. **a**, Oncoplot showing somatic events occurring in ETMRs. LOF, loss of function; WES, whole-exome sequencing; WGS, whole-genome sequencing. **b**, Overview of identified *DICER1* mutations. Alternating blue colours represent exons, yellow bars represent domains and pins represent the different aberrations found. **c**, Quantification of miRNA processing using the median ratio between 3p and 5p miRNAs ($n = 375$); each bar represents one tumour; the blue bar denotes an ETMR with a *DICER1* mutation. P values were calculated using a one-sided Mann–Whitney U -test; *** $P < 0.0005$.

fusion partners to *TTYH1* were not unique for cases without the C19MC amplification, as mate pair sequencing also identified an alternative fusion partner, *MIRLET7BHG*, in ET20 with C19MC amplification.

Finally, we identified several patients throughout the cohort with severe chromosomal instability, including four individuals without C19MC amplification (ET57, ET72, ET74 and ET104; Extended Data Fig. 5e). Taken together, our data show that ETMRs have very few recurrently mutated genes, and are largely characterized by mutually exclusive aberrations that affect the miRNA pathway, including amplifications of the C19MC and *miR-17-92* clusters and mutations in the miRNA-processing gene *DICER1*.

In contrast to SNVs and small insertions or deletions, many copy-number aberrations (CNAs) were recurrent throughout the cohort based on methylation array-derived copy-number profiling ($n = 193$). As shown previously, 90% of all ETMRs had C19MC amplification and chromosome 2 gain was detected in 76% of the cases. Other recurrent CNAs included chromosome 6q loss (25%), and gains of chromosomes 1q (26%), 17q (11%), 7 (12%), 3q (11%) and 11q (11%). However, no significant differences were found between ETMRs with or without C19MC amplification (Extended Data Fig. 6a–c). We also compared copy-number profiles of 18 matched primary relapse pairs, identifying frequently acquired CNAs, including 6q loss (22%), and gains of chromosomes 1q (33%), 17q (33%) and 7 (17%) (Extended Data Fig. 6d and Supplementary Table 7). In addition, polyploidy was detected in 18% of the primary tumours, and acquired in 28% of the patients upon relapse (Extended Data Fig. 6d, e).

Furthermore, ETMRs sequenced by whole-genome sequencing were investigated to detect translocations, inversions, deletions and insertions (Supplementary Table 8). Many clustered breakpoints that resemble chromothripsis were recurrently found surrounding C19MC, although other chromosomes were also sporadically affected

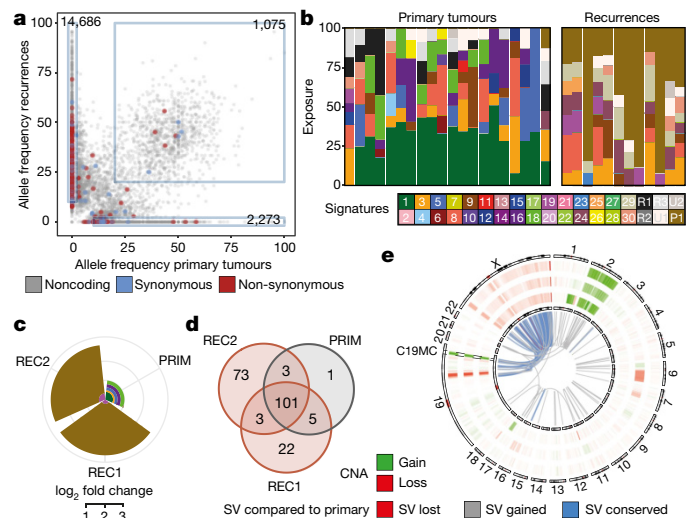


Fig. 3 | Comparison between primary and relapse tumours reveals poor conservation of SNVs but high conservation of SVs. **a**, Allele frequencies of combined SNVs found between four primary tumours and the matched initial relapses. Boxes represent SNVs that are gained, conserved and lost upon relapse. **b**, Analysis of mutational signatures of primary and relapsed tumours based on previously defined mutational signatures^{13,14}. **c**, Radial plot depicting \log_2 -transformed fold changes of the somatic SNV burden between primary tumour (PRIM) and subsequent relapses (REC1 and REC2) coloured by exposures shown in **b**. Number of mutations in the primary tumour was set to 1 for visualization purposes. **d**, Venn diagram showing the overlap of breakpoints between a primary tumour and matched relapses. **e**, Circular representation of the genome of SVs and CNAs in a primary tumour with the two matched recurrences shown in **d**. The circles represent, from outer to inner, the CNAs found in the primary tumour, the first recurrence and second recurrence. The middle part represents the different SVs found between primary tumours and recurrences. Chromosomes 19 and X have been enlarged.

(Extended Data Fig. 6f). These breakpoints were also observed in copy-number profiles: 16% of the cohort had visible alternating copy-number states around C19MC. Together, the high number of CNAs, recurrent polyploidy and recurrent complex rearrangements suggest that ETMR genomes are structurally instable.

Primary tumour–relapse comparisons

To investigate which events are retained in relapse samples and therefore potentially tumour driving, we used whole-genome sequencing data of nine matched recurrences and analysed allele frequencies of all detected SNVs in primary tumour–relapse pairs (Supplementary Table 9). Overall, we found that only 51% (2,138 out of 4,226) of all SNVs that are present at allele frequencies of at least 20% in the four matched primary tumours are also detectable at allele frequencies of 2% or more in the respective matching relapse sample (Fig. 3a). In addition, two out of four tumours did not have a single coding non-synonymous SNV detected in both primary tumour and first relapse, suggesting that acquiring somatic SNVs is not an early (driving) event (Extended Data Fig. 7).

Relapsed tumours have a large increase in somatic SNVs compared to primary tumours, suggesting that new mutations may have been induced by treatment (Extended Data Fig. 8a). To investigate this, we compared mutational signatures between primary tumours and recurrences and observed a shift from mainly signature 1 and 16 (both associated with ageing) in primary tumours to signature P1 in recurrences, which is a novel signature found in our paediatric pancancer cohort, which has not yet been associated with a mutagenesis process^{13,14} (Fig. 3b). To find the potential underlying mechanism, we tested for mutational strand bias in somatic SNVs. No bias was detected in

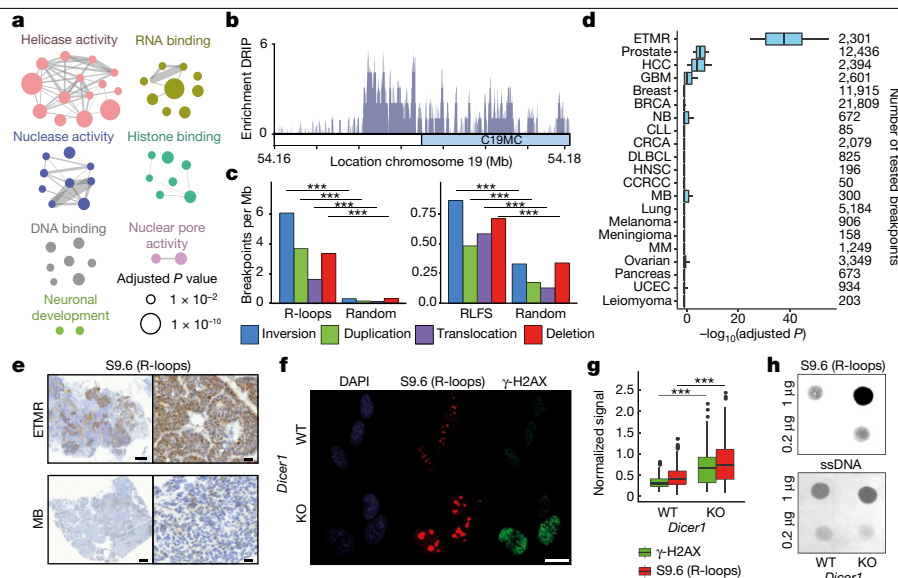


Fig. 4 | Breakpoint context reveals a possible role for R-loops in initiating ETMRs. **a**, Schematic representation of Gene Ontology (GO) term enrichment. Circles represent GO terms, sizes enrichment and colours groups based on similarity scores (co-occurrence association score >0.05) (Supplementary Table 3). **b**, Enrichment in the DRIP signal around C19MC (fold enrichment over input). **c**, Density of ETMR breakpoints ($n = 2,301$) overlapping DRIP peaks (left; $n = 16,002$) and RLFS (right; $n = 85534$) compared to random regions of the same size. P values were calculated using a two-sided χ^2 test; *** $P < 0.0005$. **d**, Enrichment of breakpoints overlapping RLFS compared to 10,000 randomly generated sets of regions of the same size for ETMRs and other entities²². Definitions of the groups are provided in Supplementary Table 10. Boxes show the range (median, first and third quartile) of Benjamini–Hochberg-adjusted P values calculated using one-sided Fisher’s exact tests, whiskers extend to upper and lower limits of the data (90% and 10% respectively).

primary tumours, however, recurrences showed an increased bias of mainly C > A, C > T and T > A mutations towards the transcribed strand (Extended Data Fig. 8b), which has previously also been observed in tumours treated with platinum-based agents and cyclophosphamide¹⁵. Comparing matched primary tumours with multiple recurrences confirmed that the composition of SNVs changed after treatment, but remained similar in subsequent relapses (Fig. 3c). Furthermore, the trinucleotide counts changed towards a signature that is highly similar to signature P1 and a cisplatin-induced mutational signature derived from cisplatin-treated cell lines¹⁶ (Extended Data Fig. 8c, d), suggesting that the majority of acquired mutations in relapsed ETMRs is induced by cisplatin treatment.

Despite the poor conservation of SNVs, there was a strong conservation of breakpoints between primary tumours and recurrences. On average 73% (specifically, 96%, 54%, 63%, 80% for the four pairs included in Fig. 3a) of all breakpoints were conserved between primary tumours and relapses, including all breakpoints that surrounded and formed C19MC aberrations (Extended Data Fig. 7). In the same sample analysed for shifts in mutational signatures, only one out of 110 breakpoints in the primary tumour was not detected in any recurrence (Fig. 3d, e). These data suggest that the formation of structural variants (SVs) is an early event in ETMR tumorigenesis. Further supporting this observation is an increased density of conserved mutations in close proximity to breakpoints, which were enriched for C > T and C > G mutations, which have previously been described in association with chromothripsis and replication stress¹⁷ (Extended Data Fig. 9). The process of forming breakpoints continues on relapse and relapsed tumours also gain many new SVs, which may have a role in tumour progression (Extended Data Fig. 7). Therefore, even though we cannot fully exclude that the acquisition of sporadic SNVs, also in non-coding regions of the genome, might

play a part in tumour formation, the high conservation of breakpoints and recurrent chromosomal instability are more likely to be driving ETMRs than the poorly conserved SNVs.

R-loop-associated chromosomal instability

In recent years, multiple papers have described an association between the formation of R-loops—structures that form upon stalling of RNA polymerase, which results in a displaced non-template single-stranded DNA loop—and DNA damage¹⁸. R-loops, which can form following disrupted helicase activity, can either facilitate or result from collision of transcription and replication and can potentially lead to chromosomal instability^{18,19}. Notably, expression of DNA/RNA helicases and processes associated with helicase activity were highly upregulated in ETMRs, suggesting a possible role for R-loops in the formation of breakpoints (Fig. 4a).

To investigate this, we used the prediction of R-loop forming sequences (RLFS)²⁰ and performed DNA–RNA hybrid immunoprecipitation (DRIP) sequencing (DRIP-seq) on the C19MC-amplified BT183 ETMR cell line. To compare how well the datasets matched, we compared the RLFS and DRIP-seq data to published DRIP-seq data for Ewing sarcoma²¹ and observed a similar genome-wide pattern (Extended Data Fig. 10a, b). R-loops in ETMRs were mostly observed in regions that surrounded the transcription start site, CpG islands and G-quadruplex-forming repeats and were also strongly enriched in peaks observed in Ewing sarcoma and RLFS (Extended Data Fig. 10c). Overall, we observed a high density of R-loops on chromosome 19, but only in ETMRs we observed a large peak surrounding C19MC (Fig. 4b). Genome wide, many breakpoints overlapped with R-loops, and the formation of breakpoints had a stronger association with regions that are enriched in R-loops, which

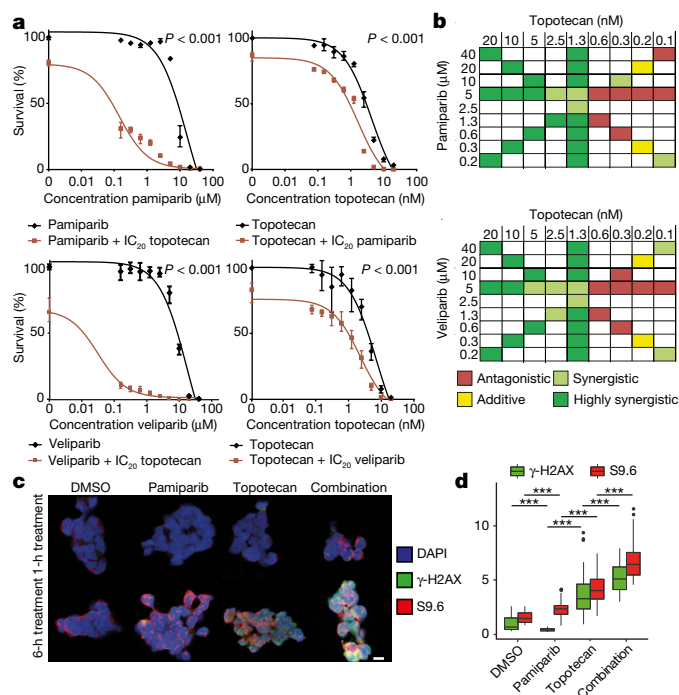


Fig. 5 | ETMR cells are sensitive to combination therapy with PARP and TOP1 inhibitors. **a**, Dose–response curves of ETMR cells treated with pamiparib or veliparib, topotecan and the combinations of drugs. *P* values were calculated using two-way analysis of variance (ANOVA). Data are mean \pm s.e.m. **b**, Calculation of synergy using the Chou–Talay method of drug treatments shown in **a**. **c**, Immunohistochemistry of ETMR cells stained with an anti- γ -H2AX antibody and the S9.6 antibody. Cells were treated using IC_{50} concentrations of each drug or combination of drugs. Scale bar, 10 μ m. **d**, Quantification of the signal shown in **c**, presented as mean signal per cell ($n=59$ DMSO-treated cells, $n=158$ pamiparib-treated cells, $n=88$ topotecan-treated cells, $n=20$ cells treated with the indicated drug combination), normalized to the total DAPI signal in the cell. *P* values were calculated using two-sided Mann–Whitney *U*-tests; ****P* < 0.0005.

was similar for breakpoints that occurred in Ewing sarcoma but not for breakpoints that occurred in other tumour entities²² (Extended Data Fig. 10d). Therefore, we compared the relative number of breakpoints in DRIP-seq peaks and RLFS compared to regions outside R-loops and found a strong enrichment of all SV types, including those that do not form the C19MC aberrations, in both DRIP-seq peaks and RLFS (Fig. 4c and Extended Data Fig. 10b). Compared to multiple other tumour entities, the enrichment of breakpoints that fell in R-loops or RLFS was specifically present in ETMRs (Fig. 4d). Furthermore, we observed an increased density of breakpoints in close proximity to R-loops and RLFS compared to other tumours that did not occur for other genomic elements (Extended Data Fig. 10e, f); these breakpoints most likely resulted from damage induced by stalled replication forks²³. Finally, we also observed an increased density of SNVs that fell in R-loops and RLFS, in line with the observed increased density of SNVs around breakpoints (Extended Data Fig. 10g). Together, these data suggest that R-loops may indeed play a part in generating breakpoints in ETMR, including the breakpoints observed around C19MC.

To validate the high levels of R-loops in ETMRs, we stained five ETMRs for R-loops using the R-loop-specific antibody S9.6 and compared the pattern of R-loop staining to staining in medulloblastomas ($n=10$) (Fig. 4e). ETMRs were highly positive for R-loops in the rosette structures whereas medulloblastomas were negative throughout the entire sections. Most ETMRs with or without C19MC amplification have miRNA-related aberrations, and multiple miRNA-processing factors have been associated with R-loop formation^{24,25}, including *DICER1*,

which has shown to be directly involved in DNA repair and contains helicase activity²⁶. To test this, we carried out dot blot and immunostaining experiments for R-loops and double-stranded breaks in a *Dicer1* knockout mouse stem cell line and its isogenic control. Indeed, both the levels of R-loops and double-stranded breaks—as indicated by increased γ -H2AX staining—were increased in the *Dicer1* knockout cells (Fig. 4f–h). In addition, sequencing detected chromothripsis events and many breakpoints in RLFS that resemble the phenotype in ETMRs around C19MC (Extended Data Fig. 10h, i). These data suggest that R-loops may cause chromosomal instability in ETMRs and defective miRNA processing might have a role in generating the increased levels of R-loops.

Efficacy of PARP and TOP1 inhibition

Finally, we investigated whether R-loops could potentially be exploited for therapy. Previously, we reported that ETMRs are highly sensitive to inhibitors of topoisomerase I (TOP1), an enzyme that can resolve R-loops among other functions^{19,27}. Other studies have shown that topotecan and irinotecan act as a TOP1 poison by covalently binding TOP1 to the DNA²⁸, which can be further enhanced by trapping PARP1, as PARP1 is able to release TOP1 by parylation²⁹. Therefore, we tested whether a combination treatment with PARP and TOP1 inhibitors would lead to a further increase in R-loops and increased response to therapy.

First, we tested whether there is synergy between PARP and TOP1 inhibition in ETMR cells. Topotecan alone is effective with half-maximum inhibitory concentration (IC_{50}) values of around 5 nM, whereas PARP inhibitors are less effective with IC_{50} values of approximately 10 μ M for pamiparib (BGB-290) and about 15 μ M for veliparib. Both PARP inhibitors, however, act highly synergistically when used in combination with topotecan and lead to a larger decrease in viability than monotherapy with any of the drugs (Fig. 5a). Synergy and cell death are only observed once concentrations of topotecan were higher than 1.25 nM, suggesting that the synergy is mostly dependent on sufficient levels of TOP1 inhibition (Fig. 5b).

Next, we investigated whether topotecan alone or combined with PARP inhibitors is correlated with R-loop levels. Topotecan treatment (6 h) increased the number of nuclear R-loops whereas pamiparib had almost no effect on R-loops (Fig. 5c). Combination therapy induced R-loops at an even higher rate and caused more DNA damage as shown by γ -H2AX staining than monotherapy. Quantification of the nuclear signals shows that γ -H2AX and R-loop signals highly correlate, suggesting that an increase in R-loops increases the amount of DNA damage (Fig. 5d). Together, these data show that PARP and topoisomerase inhibitors act synergistically in ETMRs and combination treatment could, after thorough in vivo testing, potentially be used as treatment for patients with ETMR.

Discussion

Here, we present the molecular landscape of ETMRs at diagnosis and relapse, resulting in a set of hallmarks that describe this tumour type (Fig. 6). We show that all ETMRs are molecularly similar, despite differences in histology, the presence or absence of C19MC amplification and their location in the brain. C19MC amplification is still considered the main driver based on its strong conservation and high recurrence. Moreover, we identified *DICER1* as a predisposition gene for ETMR. Our data provide further insights into the drivers of these tumours and how the tumours change after treatment, largely because of treatment with cisplatin.

miRNA expression profiles correlate strongly between ETMRs even though the underlying mechanism that deregulates miRNAs may be different. A possible explanation for this might be that tumours with C19MC or *miR-17-92* amplification oversaturate the miRNA-processing machinery. This could potentially mimic deregulated processing by *DICER1* and also explains why RNA transport—vital for miRNA

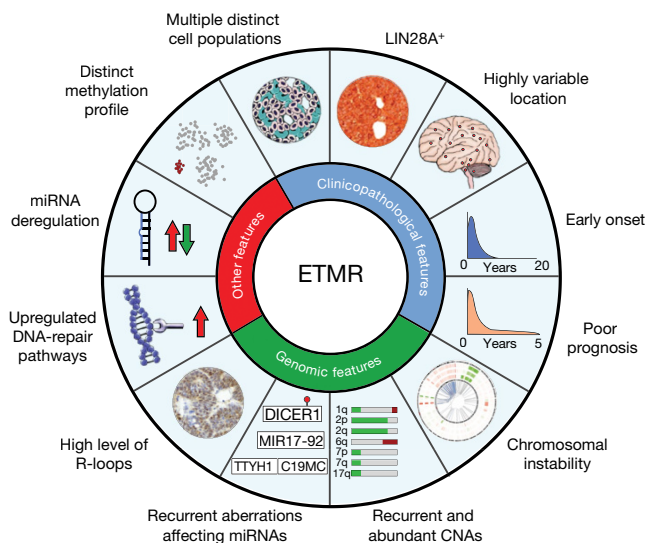


Fig. 6 | Hallmarks of ETMR. Schematic summary of ETMR characteristics.

processing and shown to be affected by miRNA saturation—is highly upregulated in ETMRs^{30,31} (Extended Data Fig. 3). Oversaturation of the miRNA machinery could also explain why all ETMRs regardless of amplification status show many structural aberrations as several members of the miRNA processing machinery, including *DICER1* and *DROSHA*, have also been associated with R-loop formation^{24,25}.

Other mechanisms that cause chromosomal instability may also play a part. Breakpoints were enriched in RLFS, which are an intrinsic property of the genome, suggesting that R-loops predispose certain loci to acquiring DNA damage. Especially because R-loops were found in close proximity of C19MC, there could be an event before C19MC amplification. Considering the early onset of the disease and strong conservation of SVs, it is possible that there is a genetic predisposition that increases the levels of R-loops, which in turn causes breakage at sites that form R-loops, leading to C19MC amplification in ETMRs. Also, C19MC amplification itself might further increase the levels of R-loops causing the number of breakpoints to increase.

One such genetic predisposition is identified here through recurrent *DICER1* germline mutations. With this finding, ETMR is added to the large spectrum of tumours that may arise in the context of *DICER1* syndrome³². The link between chromosomal instability, R-loops and *DICER1* mutations in ETMRs may occur in other cancers as well, providing a rationale for the cancer susceptibility seen in those patients. Indeed, tumours associated with *DICER1* syndrome seem to frequently acquire CNAs^{33,34}, although whether this is related to R-loop formation remains to be elucidated. We recommend that patients with ETMR and their families should consider genetic counselling for *DICER1* syndrome, at least in cases in which C19MC is not amplified.

Presence of R-loops can also have therapeutic implications. A recent study in patients with Ewing sarcoma has shown that R-loops may be associated with hyper-transcription, which leads to the sequestration of *BRCA1* and prevents *BRCA1* from promoting double-stranded break repair²¹. This explains why targeting these tumours with PARP inhibitors is successful. Here, we show that TOP1 inhibition can increase the levels of R-loops that cause an increased amount of damage to cells that already have high levels of R-loops. This mechanism may also be applicable to other drugs, providing potential novel treatment strategies.

In summary, by characterizing the events that drive ETMRs, we provide a valuable resource for future ETMR research and potential novel therapies that could be further pursued to treat patients with this deadly tumour type.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1815-x>.

- Korshunov, A. et al. Embryonal tumor with abundant neuropil and true rosettes (ETANTR), ependymoblastoma, and medulloblastoma share molecular similarity and comprise a single clinicopathological entity. *Acta Neuropathol.* **128**, 279–289 (2014).
- Pfister, S. et al. Novel genomic amplification targeting the microRNA cluster at 19q13.42 in a pediatric embryonal tumor with abundant neuropil and true rosettes. *Acta Neuropathol.* **117**, 457–464 (2009).
- Li, M. et al. Frequent amplification of a chr19q13.41 microRNA polycistron in aggressive primitive neuroectodermal brain tumors. *Cancer Cell* **16**, 533–546 (2009).
- Kleinman, C. L. et al. Fusion of *TTYH1* with the C19MC microRNA cluster drives expression of a brain-specific *DNMT3B* isoform in the embryonal brain tumor ETMR. *Nat. Genet.* **46**, 39–44 (2014).
- Eberhart, C. G., Brat, D. J., Cohen, K. J. & Burger, P. C. Pediatric neuroblastic brain tumors containing abundant neuropil and true rosettes. *Pediatr. Dev. Pathol.* **3**, 346–352 (2000).
- Sturm, D. et al. New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* **164**, 1060–1072 (2016).
- Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
- Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. & Pearl, F. M. Therapeutic opportunities within the DNA damage response. *Nat. Rev. Cancer* **15**, 166–180 (2015).
- Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
- Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
- Neumann, J. E. et al. A mouse model for embryonal tumors with multilayered rosettes uncovers the therapeutic potential of Sonic-hedgehog inhibitors. *Nat. Med.* **23**, 1191–1202 (2017).
- Anglesio, M. S. et al. Cancer-associated somatic *DICER1* hotspot mutations cause defective miRNA processing and reverse-strand expression bias to predominantly mature 3p strands through loss of 5p strand cleavage. *J. Pathol.* **229**, 400–409 (2013).
- Gröbner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
- Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Szikriszt, B. et al. A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol.* **17**, 99 (2016).
- Boot, A. et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res.* **28**, 654–665 (2018).
- Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654 (2015).
- Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597 (2015).
- El Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558 (2010).
- Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S. P. & Kuznetsov, V. A. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res.* **43**, W527–W534 (2015).
- Gorthi, A. et al. EWS-FLI1 increases transcription to cause R-loops and block *BRCA1* repair in Ewing sarcoma. *Nature* **555**, 387–391 (2018).
- Kloosterman, W. P. et al. Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. *Cell Rep.* **1**, 648–655 (2012).
- Gan, W. et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
- Lu, W. T. et al. Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. *Nat. Commun.* **9**, 532 (2018).
- Castel, S. E. et al. Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell* **159**, 572–583 (2014).
- Francia, S. et al. Site-specific Dicer and DROSHA RNA products control the DNA-damage response. *Nature* **488**, 231–235 (2012).
- Schmidt, C. et al. Preclinical drug screen reveals topotecan, actinomycin D, and volasertib as potential new therapeutic candidates for ETMR brain tumor patients. *Neuro Oncol.* **19**, 1607–1617 (2017).
- Staker, B. L. et al. The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl Acad. Sci. USA* **99**, 15387–15392 (2002).
- Das, S. K. et al. Poly(ADP-ribose) polymers regulate DNA topoisomerase I (Top1) nuclear dynamics and camptothecin sensitivity in living cells. *Nucleic Acids Res.* **44**, 8363–8375 (2016).
- Bennasser, Y. et al. Competition for XPO5 binding between Dicer mRNA, pre-miRNA and viral RNA regulates human Dicer levels. *Nat. Struct. Mol. Biol.* **18**, 323–327 (2011).
- Grimm, D. et al. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537–541 (2006).
- Schultz, K. A. P. et al. *PTEN*, *DICER1*, *FH*, and their associated tumor susceptibility syndromes: clinical features, genetics, and surveillance recommendations in childhood. *Clin. Cancer Res.* **23**, e76–e82 (2017).

33. Seki, M. et al. Biallelic *DICER1* mutations in sporadic pleuropulmonary blastoma. *Cancer Res.* **74**, 2742–2749 (2014).
34. Koelsche, C. et al. Primary intracranial spindle cell sarcoma with rhabdomyosarcoma-like features share a highly distinct methylation profile and *DICER1* mutations. *Acta Neuropathol.* **136**, 327–337 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

¹Hopp Children's Cancer Center (KITZ), Heidelberg, Germany. ²Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ³German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. ⁵Department of Cell Systems and Anatomy, University of Texas Health at San Antonio, San Antonio, TX, USA. ⁶Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio, San Antonio, TX, USA. ⁷Omics IT and Data Management Core Facility, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Department of Oncogenomics, Academic Medical Center, Amsterdam, The Netherlands. ⁹Department of Neuropathology, NN Burdenko Neurosurgical Institute, Moscow, Russia. ¹⁰Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹¹Department of Pediatrics, McGill University Health Center, McGill University, Montreal, Quebec, Canada. ¹²Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada. ¹³Department of Pediatrics, Pediatric Oncology and Hematology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ¹⁴Division of Hematology/Oncology Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. ¹⁵Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁶Department of Pediatric Oncology, Hematology and Immunology, Heidelberg University Hospital, Heidelberg, Germany. ¹⁷Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany. ¹⁸Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁹Department of Pediatric Hematology and Oncology, University Hospital Motol, Prague, Czech Republic. ²⁰Department of Pathology, St Jude Children's Research Hospital, Memphis, TN, USA. ²¹Department of Clinical Sciences Lund, Section of Neurosurgery, Faculty of Medicine, Lund University, Lund, Sweden. ²²Institute of Neurology, Medical University of Vienna, Vienna, Austria. ²³Aix-Marseille University, Neurophysiopathology Institute (INP), CNRS, Marseille, France. ²⁴Department of Pathology, APHM, Marseille, France. ²⁵Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. ²⁶Department of Pathology, Amsterdam University Medical Centers/location VUmc, Amsterdam, The Netherlands. ²⁷Department of Neuropathology, Institute of Pathology and Neuropathology, University Hospital of Tübingen, Tübingen, Germany. ²⁸Center for CNS

Tumors, Comprehensive Cancer Center Tübingen-Stuttgart, University Hospital of Tübingen, Tübingen, Germany. ²⁹Department of Pediatric Oncology, Hematology and Clinical Immunology, Medical Faculty, University Hospital Düsseldorf, Düsseldorf, Germany. ³⁰Division of Neurosurgery, Arthur and Sonia Labatt Brain Tumor Research Center, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. ³¹Pediatric Hematology and Oncology Division, University Hospital São João Alameda Hernani Monteiro, Porto, Portugal. ³²Department of Pathology, Children's Memorial Health Institute, Warsaw, Poland. ³³Institute of Neuropathology, University Hospital Münster, Münster, Germany. ³⁴2nd Department of Pediatrics, Semmelweis University, Budapest, Hungary. ³⁵Institute of Neuropathology, Brain Tumor Reference Center of the German Society of Neuropathology and Neuroanatomy, University of Bonn Medical Center, Bonn, Germany. ³⁶Department of Pathology, Toulouse University Hospital, Toulouse, France. ³⁷INSERM U1037, Cancer Research Center of Toulouse (CRCT), Toulouse, France. ³⁸INSERM U830, Laboratory of Translational Research in Pediatric Oncology, SIREDO Pediatric Oncology Center, Paris Sciences Lettres Research University, Curie Institute, Paris, France. ³⁹Pediatric Oncology Department, SIREDO Pediatric Oncology Centre, Curie Institute, Paris, France. ⁴⁰Paris Sciences et Lettres Research University, Institut Curie Hospital, Laboratory of Somatic Genetics, Paris, France. ⁴¹Department of Pathology, Montpellier University Medical Center, Montpellier, France. ⁴²Institute for Neuroscience of Montpellier (INM), INSERM U1051, Montpellier University Hospital, Montpellier, France. ⁴³Department of Pathology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁴⁴Genomics and Proteomics Core Facility, High Throughput Sequencing Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴⁵Brain Tumor Program, Texas Children's Cancer Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA. ⁴⁶Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA. ⁴⁷Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ⁴⁸Institute of Neuropathology, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany. ⁴⁹Research Institute Children's Cancer Center, Hamburg, Germany. ⁵⁰Department of Pediatric Hematology and Oncology, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany. ⁵¹Department of Pathology, NYU Langone Health, New York, NY, USA. ⁵²Division of Pediatric Hematology/Oncology, NYU Langone Medical Center, The Stephen D. Hassenfeld Children's Center for Cancer and Blood Disorders, New York, NY, USA. ⁵³Department of Radiological, Oncological and Anatomopathological Sciences, Sapienza University of Rome, Rome, Italy. ⁵⁴IRCCS Neuromed-Mediterranean Neurological Institute, Pozzilli, Italy. ⁵⁵Pediatric Glioma Research Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁵⁶Department of Pediatric Oncology/Hematology, Charité University Medicine, Berlin, Germany. ⁵⁷Department for Pediatric Hematology and Oncology, University Hospital Hamburg-Eppendorf, Hamburg, Germany. ⁵⁸Department of Pediatrics, Medical Biophysics, Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. ⁵⁹Mays Cancer Center, University of Texas Health at San Antonio, San Antonio, TX, USA. ⁶⁰These authors jointly supervised this work: Stefan M. Pfister, Andrey Korshunov, Marcel Kool. *e-mail: m.kool@dkfz.de

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Experimental model and patient details

All patient samples included in this study were acquired under informed consent according to the ICGC (www.icgc.org) guidelines or INFORM (www.dkfz.de/en/inform/) guidelines (for cases ET85 and ET86) and all included relapsed tumours from those patients. For all included cases, consent was approved by the review board of the contributing centres before shipment. Primary tumours were derived from patients who did not undergo radiotherapy or chemotherapy before surgical removal of the tumour. Material from relapses was derived from patients that received therapy as described in Supplementary Table 1 (radiotherapy or chemotherapy, which, in all cases, included cyclophosphamide and either cisplatin or carboplatin).

Samples were included based on different ETMR histologies as described previously¹, 450k/EPIC classification^{7,35} and amplification of C19MC was confirmed by visual inspection of copy number profiles. As many tumours with central nervous system primitive neuroectodermal tumour (CNS-PNET) histology could be reclassified as described previously⁶, tumours with CNS-PNET and pineoblastoma histologies were also included if there was sufficient evidence based on 450k/EPIC clustering. The presence or absence of C19MC aberrations, based on copy numbers, was validated in approximately half of the cohort of primary tumours (90 tumours) using fluorescence in situ hybridization (FISH) and/or whole-genome sequencing (WGS) and mate pair sequencing when available.

All experiments involving cell culture of ETMR cells were performed using the BT183 cell line that was obtained from J. A. Chan and the characteristics of this cell line have been described in the original publication³⁶. Cells were cultured in low-adhesion cell-culture flasks as spheroid cultures as described previously³⁶. The cell line was derived from a primary tumour of a two-year old patient and was acquired after informed consent. *Dicer1*^{-/-} and *Dicer1*^{f/f} mouse stem cells were obtained from ATCC (CRL-3220, CRL-3221) and cultured as specified by the manufacturer. All cell lines have been tested for mycoplasma contamination and were found to be negative.

Sample and library preparation for DNA sequencing

Out of the 70 primary tumours, we performed WGS for 20 cases, all with matched germline samples, and whole-exome sequencing for 17 cases without germline samples to find (recurrent) genetic events that could potentially deregulate important pathways. In addition, we performed targeted sequencing for another 33 primary ETMRs and one matched recurrence, using a gene panel of 158 genes that were found to be recurrently affected in other brain tumour types³⁷ (Supplementary Table 4).

DNA was isolated from fresh-frozen tumour material using either the Qiagen DNeasy Blood and Tissue kit or the Promega Maxwell RSC Tissue DNA Kit (AS1610). DNA from the matching blood samples was extracted using the Qiagen Blood and Cell Culture Midi Kit or Promega Maxwell RSC Blood DNA Kit (AS1400) according to the provided protocol. For samples sequenced using targeted sequencing and methylation profiling of samples for which only formalin-fixed paraffin-embedded (FFPE) material was available, DNA was extracted using the Maxwell RSC FFPE kit (AS1450) following the manufacturer's instructions.

WGS libraries were prepared using the Illumina TruSeq Nano DNA LT Library Prep or TruSeq Nano DNA HT Library Prep Kit following the manufacturer's instructions. In brief, 100 ng of genomic DNA was fragmented to approximately 350 bp using a Covaris ultrasonicator (Covaris). The fragmented DNA was then end-repaired, size-selected using magnetic beads, extended with an 'A' base on the 3' end and ligated

with TruSeq paired-end indexing adapters. The adaptor-ligated fragment libraries were enriched using eight cycles of PCR and purified 1–2 times using magnetic beads. All generated libraries were validated using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and using Agilent 2200 or 4200 TapeStation (Agilent Technologies). Libraries were sequenced on the Illumina HiSeq X (2 × 151-bp paired-end reads) according to the manufacturer's protocol.

For samples ET1, ET12, ET16, ET19, ET20, ET30, ET31, ET66, ET68, ET71, ET72 and ET87 next-generation sequencing (NGS) libraries were created as described previously using Illumina v.2 protocols³⁸. Libraries for those samples were subsequently sequenced using HiSeq 2000 v.4 instruments (paired-end 125-bp reads) using three lanes on the machine and achieving a comparable coverage to samples sequenced on the HiSeq X. The average sequencing coverage was 33.95× for tumour material and 34.99× for blood.

All samples sequenced for WGS were submitted to the DKFZ Genomics and Proteomics Core Facility and were only included for library preparation after passing all standard quality controls. Only samples with a DNA integrity over 7 were included in the study.

Sequencing of samples using whole-exome sequencing was performed by creating libraries using the Illumina TruSeq exome enrichment kit following the manufacturer's instructions after size selection. Size selection was performed by fractionation using a Covaris ultrasonicator and subsequent selection was performed using a 1.5% gel Pippin Prep cassette (Sage Science). Sequencing was performed at the Genome Quebec Innovation Centre using Illumina HiSeq 2000 instruments or at the ICGex NGS platform of the Institut Curie using HiSeq 2500 instruments for the two *DICER1*-mutated samples that were reported previously³⁹.

Targeted sequencing was performed by creating libraries using the Agilent SureSelect XT technology. Libraries were sequenced using molecular barcode-indexed ligation-based sequencing using a NextSeq500 (Illumina) instrument¹⁹. For the targeted sequencing, we sequenced genes at an average coverage of roughly 100× for the genes listed in Supplementary Table 4.

DNA methylation array

DNA methylation profiling was performed as described previously^{6,7}. DNA was extracted in the same manner as described for WGS using 500 ng as input material for fresh-frozen tissue and 250 ng input material for FFPE tissue. Array data were created using the Infinium HumanMethylation450 BeadChip array (450k array) according to the manufacturer's instructions (Illumina) at the DKFZ Genomics and Proteomics Core Facility. For a subset of samples (described in Supplementary Table 1) Methylation BeadChip (EPIC) arrays were used. CpG probes that were used for the analysis were filtered based on the presence of a common SNP within five bases of the probe, reads not mapping uniquely to the reference genome, probes mapping to the X and Y chromosome and reads not overlapping between 450K arrays and EPIC arrays.

Clustering was performed after correction of samples for the origin of the DNA (FFPE or fresh frozen) using surrogate variable analysis (sva)⁴⁰ and only the 10,000 most variable probes based on the full dataset after correction were selected for clustering. Distance was calculated using 1 – Pearson correlation and linkage was calculated using average as measure. Subsequently, *t*-stochastic neighbourhood embedding (*t*-SNE) analysis using RTSNE (v.0.13) was applied to generate the figures⁴¹.

Copy-number profiling and analysis

Copy-number profiles were created using the conumee package (v.1.3.0). For the analysis of copy numbers, the average change in copy number observed for chromosome 2 gains in annotated samples from a previous study was used as a cut-off to determine gains and losses¹. Copy numbers were subsequently manually assessed to prevent false

Article

positives and false negatives due to differences in tumour cell content or ploidy. Calling of focal amplifications or deletions was performed similarly using three times the average change in copy number for chromosome 2 gains followed by manual assessment. For samples with increased ploidy CNAs were filtered out manually that were likely to be due to ploidy changes to reduce the number of false positives.

Processing and alignment of DNA sequencing

Sequenced reads using WGS or whole-exome sequencing approaches were aligned to the human genome version 19 (hg19) reference genome using BWA-mem v.0.7.8-r455 using default settings and T-O. Lanes were sorted using biobambam bamsort (v.0.0.148) and duplicate reads were removed using biobambam bammark duplicates. For data generated using the HiSeq X Picard (v.1.125) (<https://broadinstitute.github.io/picard/>) was used to filter out duplicate mapping reads. Reads were removed if the phred-scaled base quality over the length of the read was below 25. For targeted sequencing, reads were aligned to the human reference genome version 19 (hg19) using BWA-mem v.0.7.12-r1039. Duplicated reads were removed using Picard v.1.113 and reads were sorted using SAMtools v.0.1.17-r973⁴². Reads were removed if the phred-scaled base quality over the length of the read was below 25.

Somatic SNV and insertion or deletion calling

Detection of somatic SNVs and insertions or deletions (indels) in WGS data was performed using our in-house pipeline Roddy (v.1.1.73). The pipeline is based on SAMtools (v.0.1.17-r973) mpileup and bcftools using parameter adjustments allowing for SNV calling³⁸. In addition to those filtering steps, we applied additional filtering to remove low-quality SNVs as described previously⁴³. Variants were excluded that fell into ENCODE DAC blacklisted regions, Duke excluded regions, the hiSeqDepthTopPt1Pct track from UCSC genome browser or variants that had both the reference and altered allele annotated in dbSNP (<https://www.ncbi.nlm.nih.gov/SNP/>, release 2 November 2015). In addition, we applied filtering criteria restricting the overlap of variants with features including tandem repeats, simple repeats, low complexity, satellite repeats or segmental duplications to a maximum of two. Finally, all variants were filtered out that did not fulfil the heuristic criteria of having at least five reads in the sequenced tumour covering the position, a coverage of more than 10 reads but less than 300 reads in the germline control, more than 3% of the reads of the alternative variant in the germline, no bases other than wild type or variant at the position, at least one read on every strand supporting the variant (or more than five reads in total) and a variant allele frequency at least over 10%. The final results were annotated using ANNOVAR (release 1 February 2016)⁴⁴ using data derived from Interpro (<https://www.ebi.ac.uk/interpro/>, release 15 October 2015), the Combined Annotation Dependent Depletion (CADD) phred score⁴⁵ (ljb26, release 25 September 2014), cosmic (<https://cancer.sanger.ac.uk/cosmic>, release 11 September 2014), dbSNP and clinvar (<https://www.ncbi.nlm.nih.gov/clinvar/>, release 2 March 2016).

For targeted sequencing, SNVs were detected using SAMtools mpileup and bcftools using a similar method as described for WGS with a couple of changes to increase stringency: variants were filtered out that did not have a minimum read depth of 20, a minimum root-mean-square mapping quality of 30, a minimum of three reads supporting the variant, at least one read on every strand that supported the variant and an allele frequency of at least 10%. In addition, variants were filtered out that occurred in more than 0.1% of the 1000 Genomes Project population (<http://www.internationalgenome.org>, release 24 August 2015) or in more than 0.1% of the nonTCGA ExAC population (<http://exac.broadinstitute.org>, release 23 April 2016). All reported coding SNVs and coding indels detected using panel sequencing were annotated using ANNOVAR and benign variants were filtered out.

For indel calling (<50 bp) in targeted sequencing both SAMtools mpileup and Platypus (v.0.8.1) were used as described previously³⁸.

The same criteria were applied to indels that were applied to SNVs including read depth, mapping quality, reads supporting the variant allele frequency, allele frequency in general population, filtering of benign variants and manual reviewing. Indel calling and SNV calling in exome sequencing was performed in a similar manner as targeted sequencing with an extra filtering step at alignment: as two germlines were available both were combined to form a pseudo-germline and this pseudo-germline was used for filtering variants.

Germline SNV and indel calling

Germline SNV and indel calling was performed as described previously using freebayes (<https://github.com/ekg/freebayes>, v.1.1.0) applying the same settings and using the same filtering criteria⁴⁶ using a panel of genes described in Supplementary Table 4. Genes not in the panel were excluded from analysis. In brief, raw variant predictions were filtered (QUAL >20, QUAL/AO >2, SAF >1, SAR >1, RPR >1 and RPL >1), and normalized across patients with vt (<https://genome.sph.umich.edu/wiki/Vt>, version 0.5)).

Putative germline mutations were defined as frameshift, stop gain/loss, start loss, canonical splice site, exon/gene deletions, and damaging non-canonical splice site variants or pathogenic missense variants based on ClinVar annotations. In addition, homozygous missense variants, or missense variants having a somatic second hit that are present in a functional domain with a CADD phred score higher than 20 were annotated as putative germline mutations.

Putative germline mutations were filtered out if the minor allele frequency was higher than 0.1% in any continental population based on the non-TCGA ExAC database, the 1000 Genomes Project or ESP (<http://evs.gs.washington.edu/EVS/>).

After annotation, filtered germline mutations were excluded from the analysis if annotated as benign in ClinVar. In addition, variants with a reduction in allele frequency to less than 40% in the tumour were excluded. Finally, we only retained variants that were recurrent within the cohort. The context of all remaining variants was reviewed manually using IGV to prevent false positives.

For investigation of non-coding regions, we included regions 100 bp preceding the transcription start site (based on known protein-coding genes defined in Gencode v.19) and putative enhancers (based on H3K27ac data from cultured H9 neural progenitor cells, Encode, ENCSR449AXO, Bernstein laboratory)⁴⁷.

SV discovery using paired-end sequencing data

For SV discovery, aligned reads were processed with Delly (v.0.7.5) using paired-end mapping and split-read analysis²¹. Centromeric and telomeric regions of hg19 were excluded from the analysis and, to maximize specificity for the latter, somatic filtering was run jointly on each pair of tumour and matched control samples. For germline SV discovery, we merged all SVs using the Delly merge command and genotyped all SV sites in all control samples. We then discarded SVs with an allele count of zero and filtered SVs using the germline mode of the Delly filter command. All germline SVs were then intersected with genes and protein-truncating variants, such as the *DICER1* germline exon 19 deletion, were manually inspected using IGV.

For somatic SV discovery, we first filtered each tumour sample against the matched control using the somatic mode of the Delly filter command. We then aggregated all somatic variants and genotyped them once more across all samples to fetch putative mismapping artefacts of high allele count. All somatic SVs that passed the second round of filtering against a panel of normal samples were collected for each tumour sample and overlaid on read-depth plots for manual inspection.

Mate pair sequencing

Mate pair DNA library preparation was carried out using the Illumina MP v.2 reagents and protocol as described previously³⁸. In brief, fragmentation of genomic DNA was performed using a Hydroshear

device to an insert size of 4.5 kb followed by sequencing with Illumina HiSeq 2000 instruments. Alignment was performed using Eland (v.2) retaining only uniquely aligned reads for downstream rearrangement analysis using Delly. Mate pair sequencing was applied on complex rearrangements detected using WGS mainly as validation and reconstruction of the rearrangements but was not applied for discovery of structural alterations.

Oncoplots

Oncoplots were generated using custom scripts. Events (either germline or somatic) were only included if the minor allele frequency in the non-TCGA ExAC population was under 1% or unknown and the variant allele frequency was higher than 20%. In addition, every event had to fulfil one of the following criteria: the gene was either recurrently affected with at least one event present in the WGS cohort or a loss-of-function mutation within a recurrently affected pathway. Events were only included in the figure if events occurred in ETMRs without C19MC amplification, in multiple ETMRs with C19MC amplification or were loss-of-function mutations in deregulated pathways. For the CNA oncoplot, all CNAs were included after manual filtering for CNAs that were probably the result of increased ploidy.

Conservation between primary and relapse tumours

Conservation of events between primary tumours and recurrences was performed using somatic calls of SNVs, indels and SVs. Variant allele frequencies were recalculated at the defined positions using SAMtools mpileup followed by Varscan pileup2cns (v.2.3.9)⁴⁸. Variants were only included if the coverage in both germline and tumour was at least 10 reads, the coverage in both primary and recurrence was at least 10 reads, and no alleles other than wild type or variant were detected in the germline, primary tumour or recurrence. In all sample combinations, more than 95% of SNVs had sufficient reads for further analysis. Retained events were defined as events that had an allele frequency over 20% in both primary and recurrence, gained mutations in recurrences were defined as mutations that had an allele frequency less than 2% in the primary tumour and more than 10% in the recurrence and events that were lost in recurrence were defined as events with allele frequencies over 10% in primary tumours but less than 2% in recurrences.

Calculation of mutational signatures

Mutational signatures were calculated as described previously and compared to 30 previously published signatures¹⁴ and the recently described P1 signature¹³. In brief, somatic mutations and mutation context (adjacent bases) were extracted from the called SNVs and used to construct a catalogue of every possible context for all samples. We calculated the exposures using the formula $M = P \times E$ where M is the mutational catalogue; P is the mutational signature and E is exposures (contribution of mutational process within the mutational landscape).

Signatures were calculated for both primary tumours and recurrences with a mutation count over 200. Whole exomes and samples sequenced using targeted sequencing were therefore excluded from analysis. Exposures present under 5% in a sample were removed and exposures over 5% were reconstructed to form 100% of the sample to reduce the amount of noise. Changes in mutational signatures between primary tumours and recurrences were also calculated using only the 30 previously published signatures¹⁴ (data not shown). Signature calculation was also performed de novo and five signatures were called. Compared to the 30 previously published signatures, one novel signature was found. This novel signature had a cosine similarity over 0.85 only for signature P1, therefore the P1 signature was included in the analysis instead of only the 30 classical signatures.

Calculation of transcriptional asymmetry

Transcriptional asymmetry was calculated using the MutationalPatterns (v.1.4.3) R package based on previously described methods^{49,50}.

Mutations were considered that overlapped gene bodies and were combined from either all primary tumours or all relapses. Substitutions were assigned as transcribed when present on the same strand as the gene definition (UCSC, hg19) and untranscribed if present on the opposite strand compared to the gene definition. When multiple genes in different directions overlapped, events were excluded in this region. Events were split in different substitution types and asymmetry was calculated separately for every substitution type for both primary and relapse tumour samples.

RNA profiling

RNA was isolated from 28 ETMRs using the Qiagen RNeasy Mini kit or the Maxwell RSC simplyRNA Tissue Kit (AS1340) using homogenized tumour tissue. All RNA isolations were performed according to the manufacturer's protocol. Quality control was performed using an Agilent Bioanalyzer 2100 instrument and samples with an RNA integrity index over 7 were included. Samples were profiled using Affymetrix GeneChip Human Genome U133 Plus2.0 arrays at the DKFZ Genomics and Proteomics Core Facility applying the manufacturer's protocol for preparation, hybridization and quality control.

Analysis of mRNA data

Samples were normalized and probe detection P values were calculated using the MASS.0 algorithm (Affymetrix), followed by quality control of the percentage of present calls and manual inspection of *GAPDH* levels and 5' to 3' ratios. All probes were filtered to contain every gene once in the analysis. Analysis of atypical teratoid rhabdoid tumours (ATRT)⁵¹, medulloblastoma (MB)⁵² and CNS-PNET⁶ data was performed in the same way for every sample included in the analysis.

Differential expression was calculated using a two-way ANOVA using a P value of 0.01 as a cut-off (corrected using false-discovery rate (FDR)). In addition, an absolute fold change of 2 and a minimum \log_2 -transformed gene expression value of 5 in at least one of the compared sets were used as criteria for the definition of differentially expressed genes. Heat maps were generated using supervised clustering of expression of 450 DNA repair genes shown in (Supplementary Table 4) and pathways described previously⁸. All samples were z-score normalized and clustering was performed using Euclidean distance. All mRNA expression data were analysed using the R2 platform (<http://R2.amc.nl>).

t -SNE clustering was performed by using z-score-normalized gene expression using only representative probes by applying HugoOnce (<http://R2.amc.nl>). Gene expression data were derived from a set of 580 samples deposited in the R2 platform and 28 ETMRs, which were all processed in the same way. Using these samples, a 50-dimension principal component analysis was performed followed by t -SNE clustering using RTSNE.

GO and KEGG enrichment analysis

GO-term enrichment and KEGG pathway enrichment was performed using TOPPgene⁵³ and filtered using significant terms also found using the DAVID algorithm (v.6.8)⁵⁴. For all terms a P value cut-off of 0.05 was used and P values were corrected using FDR, Bonferroni and Benjamini–Hochberg corrections. For GO term enrichment a gene limit of 200 was chosen to obtain more specificity in the results. Processes were defined using similarity between GO terms that was calculated using NaviGO⁵⁵ using a minimum co-occurrence score of 0.05. For terms without similarity to any significant terms this was manually assessed.

Estimating cell populations from bulk mRNA sequencing

The Cibersort algorithm (v.1.06)⁹ was applied to estimate the relative frequencies of different cell populations using normalized expression data. Signature genes were derived from the median gene expression levels of defined cell populations from the prefrontal cortex as described previously¹⁰. Only informative genes were selected that had a median expression higher than 1 in any of the subgroups. Both relative

Article

and absolute mode were applied using 500 permutations to estimate the relative abundance of each sample type.

Microdissection of FFPE material

As a representative FFPE tissue sample, ET174 was histologically identified, targeted and microdissected with a puncher for nucleic acid extraction. DNA was extracted in a similar manner as described for other FFPE material, RNA was extracted using the automated Maxwell system with the Maxwell 16 LEV RNA FFPE Kit (Promega), according to the manufacturer's instructions. To evaluate FFPE RNA quality, we used the percentage of RNA fragments >200 nt fragment determination value (DV₂₀₀). Only RNA samples with DV₂₀₀ > 70% were included for sequencing on a NextSeq 500 (Illumina). RNA-sequencing data were quantified using the quant option of kallisto (v.0.43.0) using standard settings⁵⁶.

miRNA sequencing and processing

Small RNAs were isolated as described previously^{57,58} from fresh-frozen tumour material. In brief, total RNA was extracted using guanidinium isothiocyanate/phenol extraction followed by 3'-adaptor ligation of barcoded adenylated adaptors. Samples were pooled in two sets of five samples. Subsequently, gel electrophoresis was used to isolate small RNAs (19–35 nt) and purified using ethanol precipitation. Fragments were then amplified using standard PCR, isolated using gel electrophoresis and purified using ethanol precipitation. Samples were sequenced on a HiSeq 2000 v.4 machine.

Small RNA-sequencing data were aligned using Bowtie v.1.00 (--seed-mms 1 --maqerr 1000 --seedlen 21 --norc -M1 --best --strata) to mirbase 18 after reads were selected by cutting off adapters using cutadapt v.9.5 (-e 0.1 -m 18 -M 34 -O 8 TCGTATGCCGTCTTCTGCTTG) and taking reads between 18 and 34 nucleotides. Subsequently, reads were counted using SAMtools 0.1.17-r973 mpileup.

Analysis of miRNA data

Mature miRNA counts were quantified relatively to the total read count separately for both the 5p and 3p strand and the resulting reads per million (RPM) values were used in all subsequent analysis. Supervised clustering of miRNAs was performed using only significantly differentially expressed miRNAs between ETMRs with C19MC amplification and other samples (excluding ETMRs without C19MC amplification) using adjusted negative binomial testing with the DESeq2 package (v.1.18.1)⁵⁹. Samples were normalized using z-score normalization and clustered using hierarchical clustering using average as method. Unsupervised clustering was performed after filtering for miRNAs that had an expression over 32 RPM in any included sample. Samples were normalized and clustered in the same way as the supervised analysis.

miRNA processing was quantified by comparing the normalized levels of 3p mature miRNAs to the levels of 5p mature miRNAs using sequences provided by miRbase 18⁶⁰. For every sample, RPM levels were compared for each 5p versus 3p miRNA and the mean value was calculated as the processing ratio. miRNAs with multiple 5p or 3p variants were averaged. All miRNA data were compared against a set of 10 random samples per entity derived from unpublished ICGC data (C. Aichmüller et al., manuscript in preparation) that were processed in the same way as the ETMR data.

Analysis of R-loop levels

DNA–RNA hybrids were extracted from tissue derived from ETMR patient-derived xenograft (PDX) models (BT183) that were treated using topotecan or saline as described previously²⁷. Tumours were subsequently frozen and pelleted using ultracentrifugation. DNA–RNA hybrids were extracted as described previously using the same protocol that is applied for cultured cells²¹. DNA was extracted using proteinase K followed by phenol–chloroform extraction and ethanol precipitation. Subsequently the DNA was fragmented using the restriction enzymes

HindIII, EcoRI, BsrGI, XbaI and SspI (New England Biolabs). Digested DNA was subsequently incubated with the anti-DNA–RNA hybrid antibody S9.6 (Merck, MABE1095) and immunoprecipitated using agarose beads. Bound DNA–RNA hybrids were eluted and incubated with proteinase K and cleaned with an additional phenol–chloroform–ethanol extraction. The DNA was subsequently sonicated and sequenced using a HiSeq 2000 machine with a 50-bp single-read protocol. Each treatment condition was performed in duplicate and both RNase H and the input was included as negative controls.

DRIP-seq analysis

DRIP-seq data were aligned to hg19 using BWA using standard settings. Aligned reads were normalized to the input signal using bamCompare (deeptools 3.0.2) using log₂ fold change increase as output for visualization purposes⁶¹. Peak calling was performed using MACS v.2.0 (<https://github.com/taoliu/MACS>) using the settings -g 2.7e9 -q 0.05, -B -m 2 30 and the input signal as background. Peak files were combined for each condition and each subsequent analysis was performed for both topotecan-treated samples and untreated samples. Peak calling was compared to data from Ewing sarcoma²¹ and correlated to mRNA expression to ensure the quality of the dataset.

RLFS were predicted using the qmRLFS finder (v.1.5) tool²⁰ by applying standard settings and using models m1 and m2 for both the human genome (hg19) and mouse genome (mm10) depending on the analysis. Both the R-loop initiation zone and R-loop elongation zone were included in RLFS peaks.

Calculating R-loop enrichments

For the analysis comparing the overlap of SVs with R-loops, overlapping SVs were defined as either the start position of the SV, the end position of the SV or both positions falling in either an RLFS or DRIP peak depending on the analysis. Duplicate breakpoints were excluded when overlapping between matched primary tumour and relapse. The analysis comparing breakpoints in R-loops against the rest of the genome was performed by taking the total number of SVs falling in R-loop regions and by comparing this to 10,000 randomly generated regions of the same size, with the same number of regions as the R-loop regions. For the comparison against other tumour entities 68,018 somatic rearrangements were used from a previously published study²² and locations were compared to either RLFS or R-loops derived from ETMRs. *P* values were calculated using Fisher's exact tests and corrected for multiple testing using Benjamini–Hochberg correction. The same 10,000 random regions were used for each entity and a *P* value was calculated using Fisher's exact test for every iteration and adjusted using Bonferroni correction. Only entities with more than 50 breakpoints were used for the comparison. For comparisons of R-loops with genomic regions ranges from both non-B-DB 2.0⁶² and repeatmasker 4.0.8. (<http://www.repeatmasker.org/>) were aggregated and the median enrichment over the input signal was calculated for 10,000 randomly selected regions of every DNA element.

FISH analysis

FISH was applied as described previously to validate C19MC amplification⁶³. In brief, two probes corresponding to the 19q13.42 and 19q13 loci were applied, using the 19q13 probe as a reference. The minichromosome formed between chromosome 11 and 13 was validated using probes at the loci 11q22.2 and 13q31.3. Ploidy of 28 different ETMRs was validated in a similar manner using FISH probes at the chromosome 9 and chromosome 11 centromeres as those chromosomes were found to be relatively stable in our cohort.

Detection of R-loop levels

DNA–RNA hybrids were extracted from cell pellets of *Dicer1* wild-type and knockout mouse cells after lysis for 12–24 h, with a solution containing TE, SDS and proteinase K as described previously⁶⁴. DNA–RNA

hybrids were subsequently extracted using phenol–chloroform isoamyl alcohol (Affymetrix, 75831) followed by phase-lock gel separation (Quantabio, 5prime phase lock gel light, 2302820). DNA–RNA hybrids were subsequently purified using ethanol precipitation. DNA–RNA hybrids were diluted in 10× SSC buffer and loaded on a charged nylon membrane (Roche, 11209299001). Membranes were blocked in 1× TBS containing 5% skimmed milk and incubated overnight using an anti-DNA–RNA hybrid (S9.6) antibody (Merck, MABE1095) at a concentration of 1:1,000. Subsequently an HRP-linked secondary antibody (Santa Cruz, sc-2005) was used followed by incubation with ECL (GE Healthcare, RPN2232). Methylene blue staining (Sigma-Aldrich, M9140) supplemented with 0.3 mM NaOAc was used as a loading control for all dot blots. Dot blots were developed using a chemoluminescence imaging system (Intas, ECL chemostar). Experiments involving dot blots were performed five times to ensure reproducibility.

R-loop stainings for tumour material were performed on formalin-fixed tissue slides of five ETMRs, five group 4 medulloblastomas and five WNT medulloblastomas using the same conditions as described previously²¹. For all immunohistochemistry experiments, representative photographs of the slides were taken in addition to photographs covering the entire slide.

In vitro drug response

Response to treatment with pamiparib (BGB-290) (MedChemExpress, HY-104044) or veliparib (Abbvie, S1004) and topotecan (ApexBio, B4982) was evaluated using dose–response curves. BT183 cells were plated in 96-well plates 24 h before treatment. Twofold increasing concentrations of the inhibitor were added with concentrations ranging from 0.08 nM to 20 nM for topotecan and 156.25 nM to 40 mM for pamiparib and veliparib. Cell viability was measured 72 h after treatment using an automated plate reader after using CellTiter-Glo (Promega, G7570). Graphpad Prism was used to generate dose–response curves. For combination treatment, the IC₂₀ was determined using the dose–response curves of single inhibitors and added to increasing concentrations of the other drug. Synergism between inhibitors was calculated by applying the Chou–Talay method as described previously⁶⁵.

Immunofluorescence

BT-183 cells were grown on sterilized glass coverslips coated with human laminin (Sigma, L2020) and poly-L-lysine (Sigma, P8920) for 24 h before treatment with DMSO (0.6%), topotecan, pamiparib or both pamiparib and topotecan. Before fixation using 4% PFA, spheres were washed with 0.5% NP-40 for 5 min on ice to increase the permeability. Subsequently, cells were permeabilized using 0.3% Triton X-100 for 10 min followed by 1 h blocking with 10% donkey serum. Slides were incubated with primary antibody 1:100 for γ-H2AX (Abcam, ab11174) or S9.6 in 10% donkey serum overnight. Subsequently, coverslips were incubated with DAPI (1:1,000) and Alexa Fluor 488/568 (1:500) (Life Technologies). Coverslips were mounted on glass slides using mountant (Invitrogen, P36930) and imaged using a confocal microscope (Zeiss LSM 800). Representative pictures were taken at 400× magnification and processed using Airyscan processing (included in the Zeiss LSM 800 software) and applying the middle of 21 z-stacks. Red and green channels were normalized to the blue channel (DAPI) in all shown pictures.

Signals were quantified using ImageJ, applying a custom-made macro over three random neurospheres of seven z-stacks each for every treatment condition. Signal intensity was measured in separate regions by binarizing DAPI signal and using watershed image segmentation to generate regions that simulate cells within the neurosphere. Subsequently, background subtraction (rolling ball radius of 50 pixels) was applied on every z-stack for every channel. For each channel the average signal per cell was calculated and the γ-H2AX signal and S9.6 signal was normalized to DAPI signal per cell.

Dicer1 wild-type and *Dicer1* knockout mouse cells were grown and mounted in the same way but were not grown on coated coverslips as cells were adherent. Representative images were taken at a 630× magnification using the middle of 11 z-stacks. Quantification was performed at a 200× magnification using three representative regions with more than five cells using five different z-stacks. Background correction, signal calculation and normalization was performed in the same manner as for the neurospheres.

Statistics and reproducibility

Statistical enrichment was calculated using Fisher's exact tests for calculations of binary values and proportions unless stated otherwise. For comparison between continuous values Mann–Whitney *U*-tests (also known as Wilcoxon rank-sum tests) were used unless otherwise stated in either the figure legend or Methods section. Multiple testing correction was performed using the Benjamini–Hochberg method unless otherwise stated in the figure legend. All statistical analysis was performed in R (v.3.4.3) except for dose–response curves, which were calculated using GraphPad Prism 6. Unless stated otherwise, results from representative experiments were performed at least three times. Pictures shown in Fig. 4f, g are representative of three different regions each containing more than three cells, the experiment is representative of three biological replicates. Dot blots in Fig. 4h are representative of five biological replicates. Dose–response curves in Fig. 5a were based on three biological replicates for each concentration of each drug and combination of drugs. The experiment was repeated three times and a representative curve is shown for each drug combination. Images shown in Fig. 5c, d are representative of three different regions each containing more than three cells, experiment is representative of three biological replicates.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw and processed 450K/EPIC methylation values, and raw and processed expression data for all included ETMRs are deposited at the Gene Expression Omnibus (GEO) under accession number GSE122038. All NGS data are deposited at the European Genome-phenome Archive (EGA) under accession number EGAS00001003256. Source Data for Figs. 1a–c, 2c, 3b, c, 4d, g, 5a, b, d and Extended Data Figs. 1a, 2a–g, 4b, c, h, 5c, 6b, c, 8a–d, 9b, c, e–g, 10g are provided with the paper.

Code availability

All custom code used to generate the data in this study is available upon reasonable request.

35. Hovestadt, V. et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125**, 913–916 (2013).
36. Spence, T. et al. A novel C19MC amplified cell line links Lin28/let-7 to mTOR signaling in embryonal tumor with multilayered rosettes. *Neuro Oncol.* **16**, 62–71 (2014).
37. Sahm, F. et al. Next-generation sequencing in routine brain tumor diagnostics enables an integrated diagnosis and identifies actionable targets. *Acta Neuropathol.* **131**, 903–910 (2016).
38. Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
39. Uro-Coste, E. et al. ETMR-like infantile cerebellar embryonal tumors in the extended morphologic spectrum of *DICER1*-related tumors. *Acta Neuropathol.* **137**, 175–177 (2019).
40. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
41. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
42. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

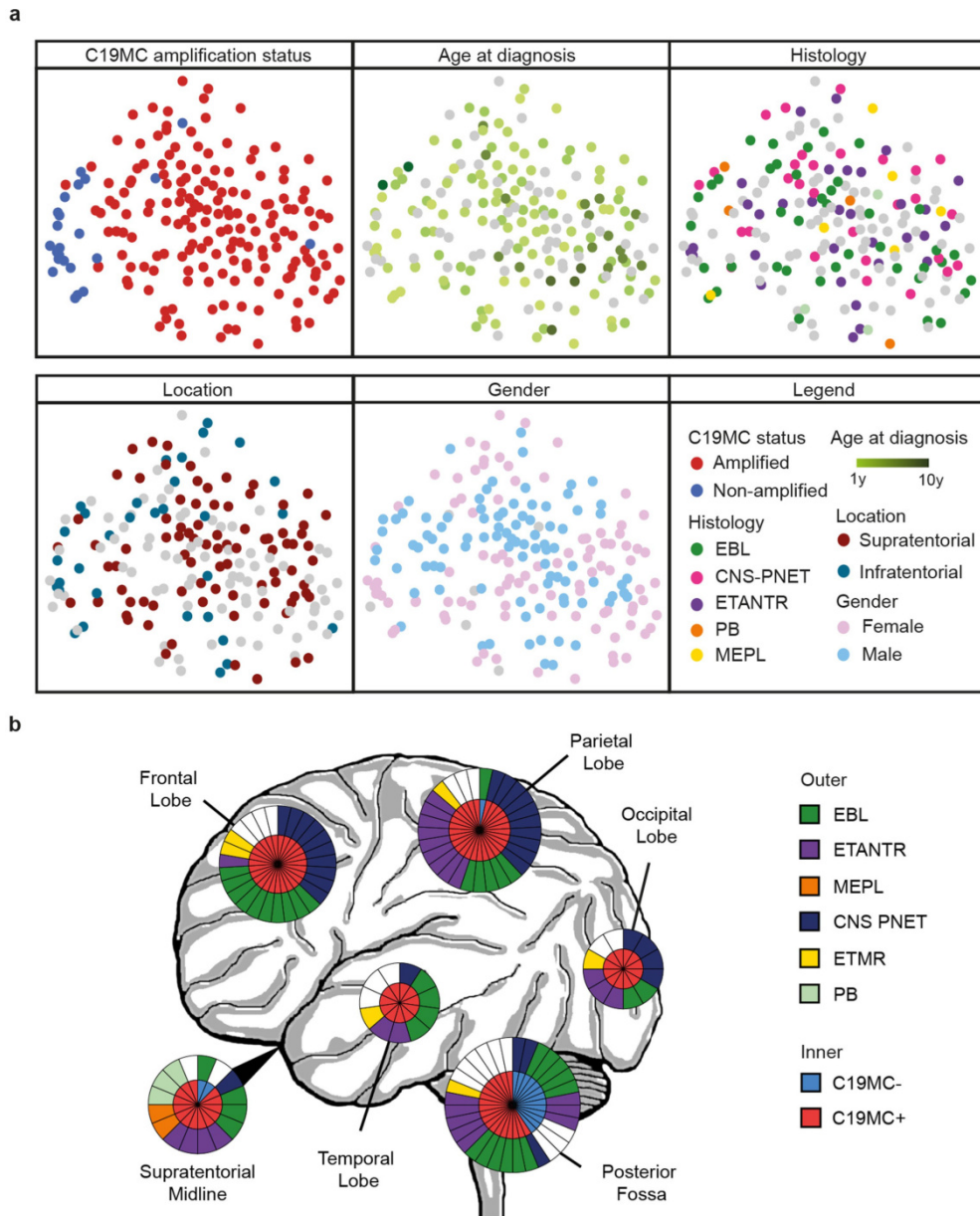
43. Kool, M. et al. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**, 393–405 (2014).
44. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
45. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
46. Waszak, S. M. et al. Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort. *Lancet Oncol.* **19**, 785–798 (2018).
47. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
48. Koboldt, D. C. et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009).
49. Haradhvala, N. J. et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell* **164**, 538–549 (2016).
50. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
51. Johann, P. D. et al. Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* **29**, 379–393 (2016).
52. Northcott, P. A. et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
53. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
54. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
55. Wei, Q., Khan, I. K., Ding, Z., Yerneni, S. & Kihara, D. NaviGO: interactive tool for visualization and functional similarity and coherence analysis with gene ontology. *BMC Bioinformatics* **18**, 177 (2017).
56. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
57. Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
58. Hafner, M. et al. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**, 3–12 (2008).
59. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
60. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
61. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
62. Cer, R. Z. et al. Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.* **41**, D94–D100 (2013).
63. Korshunov, A. et al. Focal genomic amplification at 19q13.42 comprises a powerful diagnostic marker for embryonal tumors with ependymoblastic rosettes. *Acta Neuropathol.* **120**, 253–260 (2010).
64. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).
65. Chou, T. C. Drug combination studies and their synergy quantification using the Chou–Talalay method. *Cancer Res.* **70**, 440–446 (2010).
66. Anderson, N. D. et al. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* **361**, eaam8419 (2018).

Acknowledgements We thank the DKFZ sequencing core facility for technical support, assistance with data generation and data management, the DKFZ light microscopy facility for their assistance in generating microscopy images and BeiGene for providing pamiparib. This work was supported by the ICGC PedBrain Tumor Project, funded by the German Cancer Aid (109252) and by the German Federal Ministry of Education and Research: BMBF grants 01KU1201A (PedBrain Tumor) and 01KU1505A (ICGC-DE-MINING). Additional funding was awarded by the NIH (K22ES012264, 1R15ES019128 and 1R01CA152063), Voelcker Fund Young Investigator Award and CPRIT (RP150445) to A.J.R.B.; CPRIT (RP101491), NCI T32 postdoctoral training grant (T32CA148724), NCATS TL1 (TL1TR002647) and the AACR-AstraZeneca Stimulating Therapeutic Advances through Research Training grant to A.G.; CPRIT (RP140105) to J.C.R.; and NCI (P30CA054174) to the sequencing core facility. S.L. and M.K. are supported by the Solving Kid’s Cancer foundation and the Bibi Fund for Childhood Cancer Research. A.K. is supported by the Helmholtz Association Research Grant (Germany). M. Ryzhova is supported by an RSF Research Grant (18-45-06012). J.O.K. was funded by an ERC starting grant.

Author contributions S.L. performed data analysis and interpretation. S.L., S.M.P. and M.K. wrote the manuscript with input from all co-authors. S.L., A.G., P. Landgraf, B.H., D.T.W.J., J.O.K., P. Lichter, A.H., A.J.R.B., S.M.P., A.K., S.W. and M.K. generated or contributed to the generation of sequencing data. S.L., S.N.G., T.R., S.M.W., A.G., I.B., J.K., M. Sill, V.H., D.A.Z. and S.P.-C. performed bioinformatic analyses. S.L., C.S., M.M., S.B., S.K., J.-M.H., N.M., B.S. and J.A.C. contributed to the design of experiments and conducted experiments. J.C.R., M. Ryzhova, A.J.R.B. and A.K. performed histopathological analysis on the samples. M. Ryzhova, J.A.C., T.M., B.H., O.W., J.E., F.S., D.S., D.W.E., B.A.O., A.D., C.H., D.F.-B., P.W., J.S., M. Remke, M.D.T., M.J.G.-d.-C., M.E., W.G., M.H., P.H., T.P., E.U.-C., F.B., V.R., S.A., J.M.-P., X.-N.L., U.S., M. Snuderl, M.A.K., F.G., N.J., A.v.D., K.v.H., A.H. and A.K. provided tumour samples and metadata. S.M.P., A.K. and M.K. managed the project and provided leadership.

Competing interests The authors declare no competing interests.

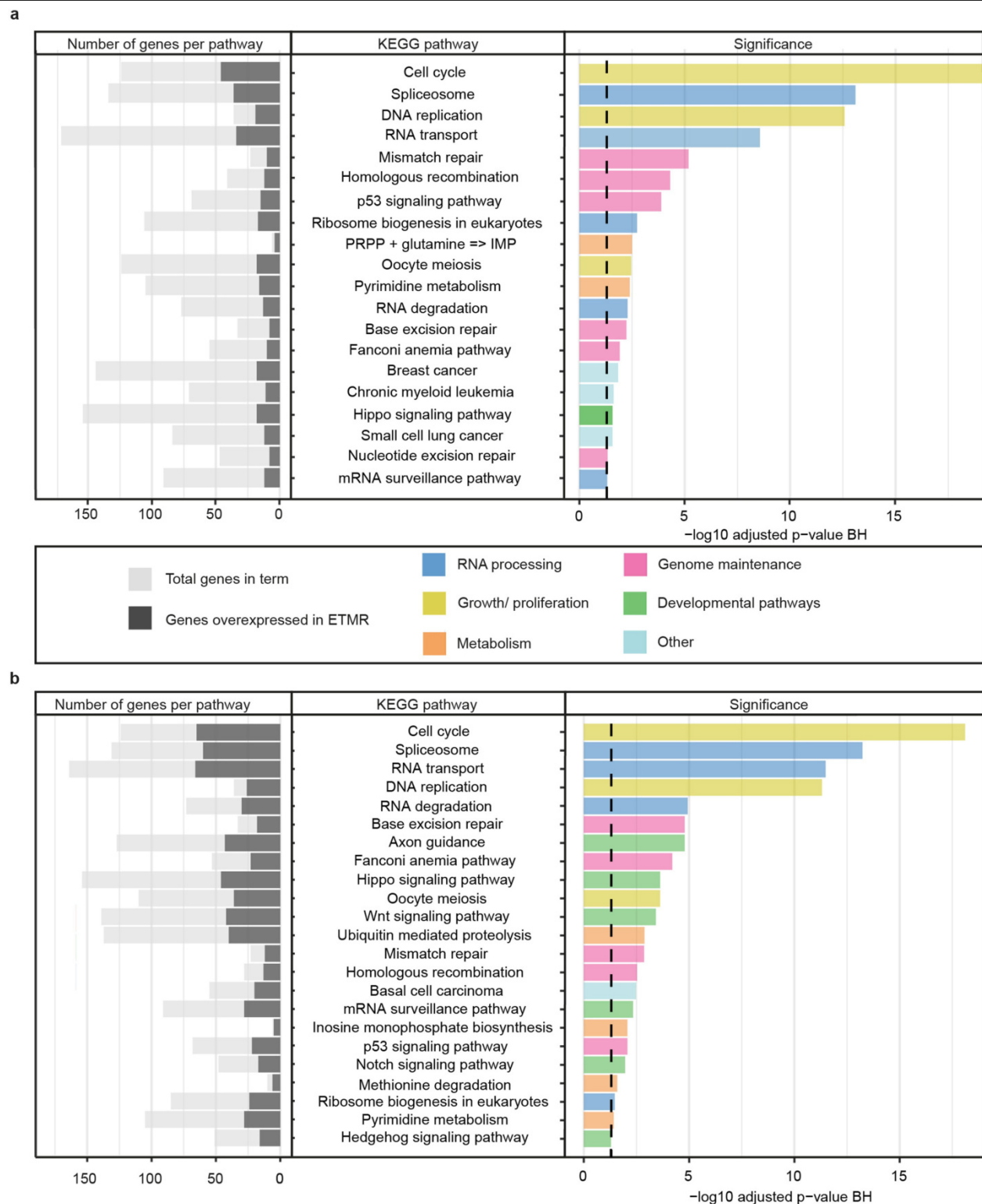
Additional information
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1815-x>.
Correspondence and requests for materials should be addressed to M.K.
Peer review information *Nature* thanks Jeffrey Chuang, Richard J. Gilbertson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Clinicopathological differences are not associated with molecular subgrouping. a, *t*-SNE clustering analysis of DNA methylation profiles of 193 ETMRs. Samples were coloured according to their clinical, histological or molecular annotation. **b**, Schematic representation of location (position of the circle), histological diagnosis (outer ring) and C19MC status

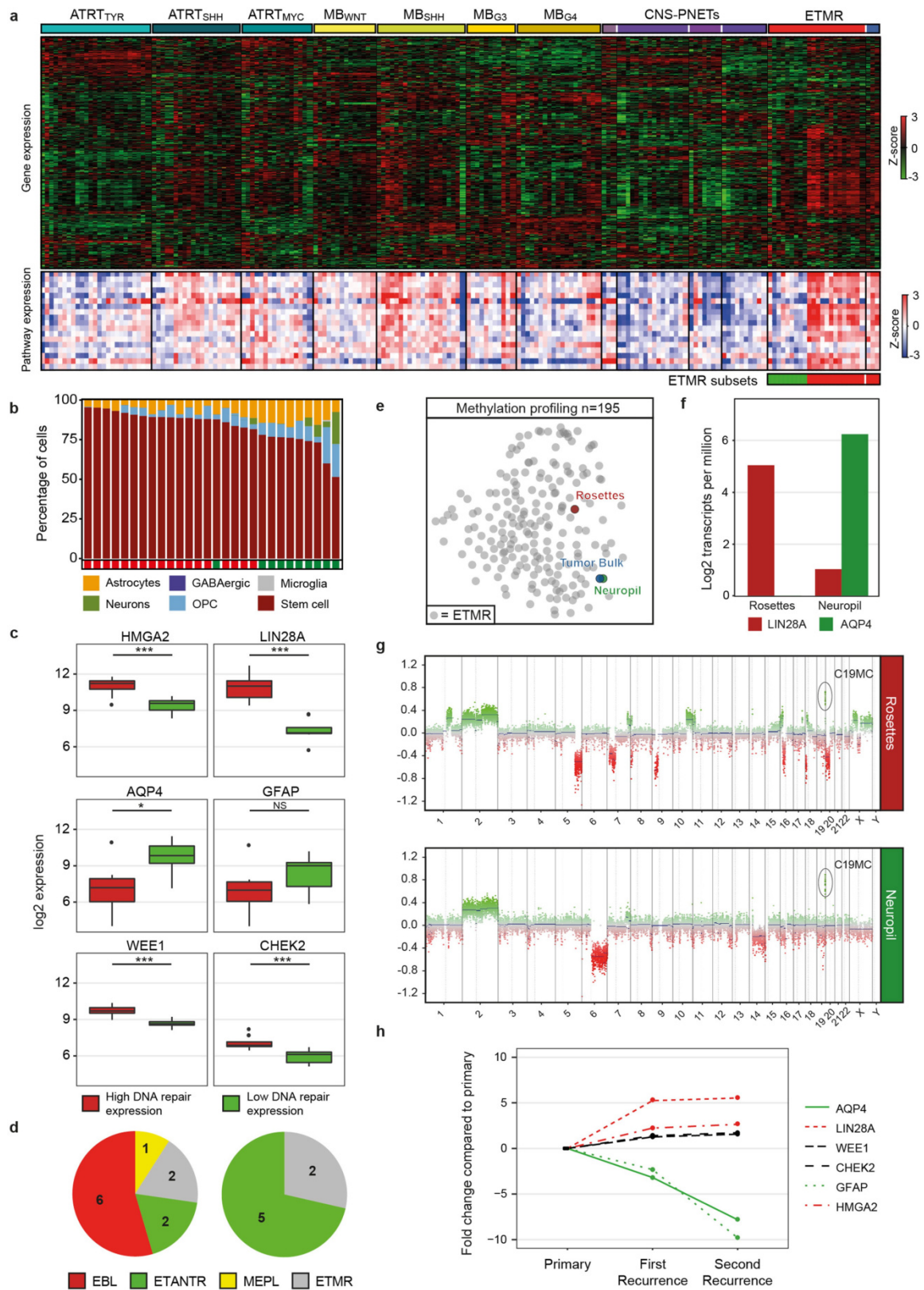
(inner ring) of ETMRs. Circle size denotes the relative number of primary tumours that have been diagnosed in each part of the brain; each wedge represents one tumour. Tumours could be assigned to multiple locations depending on the diagnosis. Tumours were excluded for which no information on the site of occurrence was available.

median expression of mature miRNAs derived from ETMRs ($n = 7$) against normal brain ($n = 8$), other entities ($n = 10$ for all entities) or ETMRs without C19MC amplification ($n = 3$). miRNAs that had a median expression below 32 RPM in either of the compared entities were excluded. miRNAs that were differentially expressed between ETMRs (with and without C19MC amplification) against other entities (two-sided negative binomial, Benjamini-Hochberg-adjusted $P < 0.05$) are highlighted. For each comparison, the Pearson correlation was calculated ($P < 0.0005$ for all comparisons).



Extended Data Fig. 3 | KEGG pathway enrichment in ETMRs. a, b, Summary of KEGG pathway enrichment of ETMRs ($n = 28$) against healthy brain tissues ($n = 38$) (a) or 580 different brain tumours (b). Pathways are coloured by

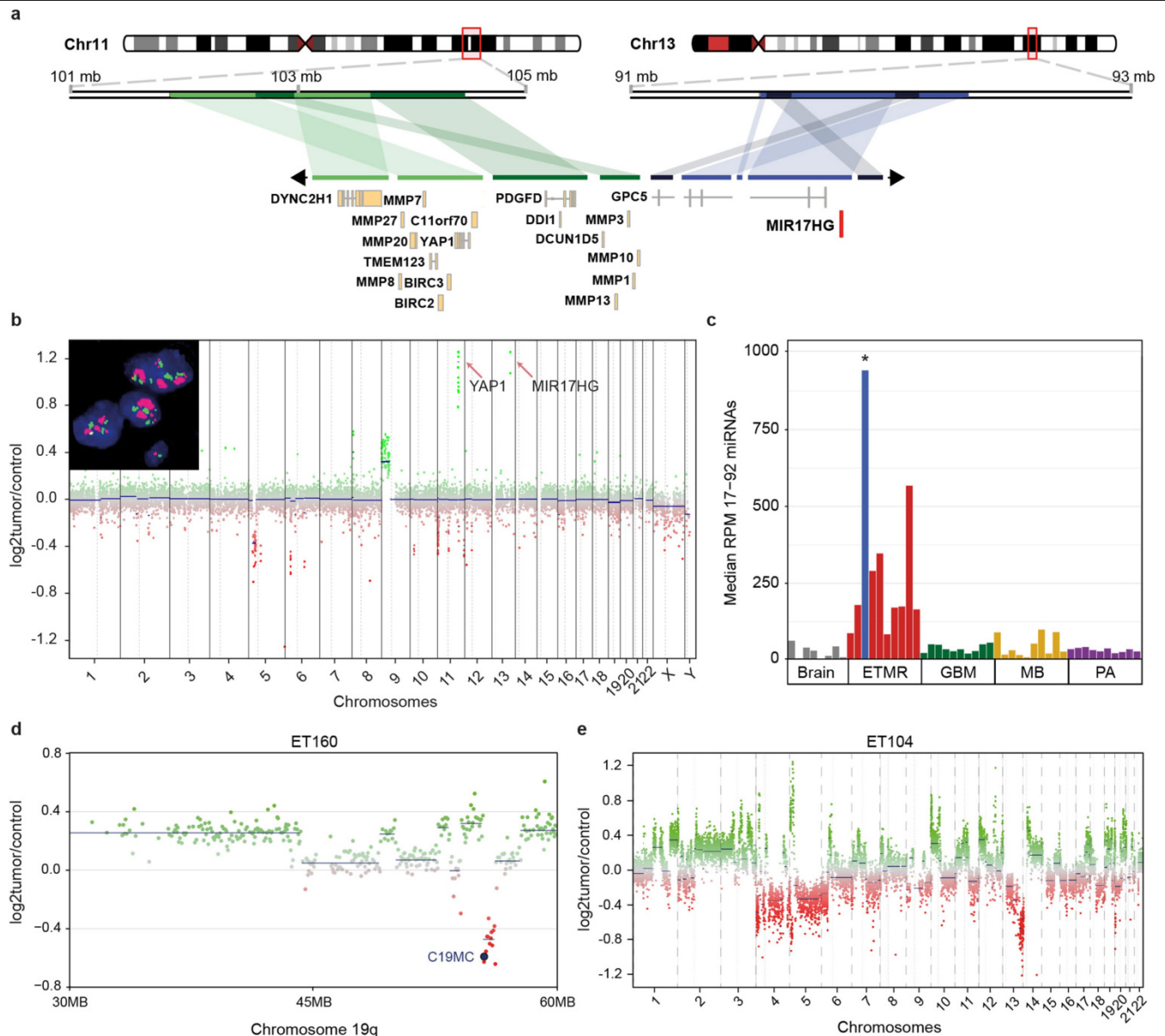
similarity based on NaviGO co-occurrence scores⁵⁵ and manual assessment. Significantly upregulated genes were calculated using ANOVA (FDR-adjusted $P < 0.01$).



Extended Data Fig. 4 | See next page for caption.

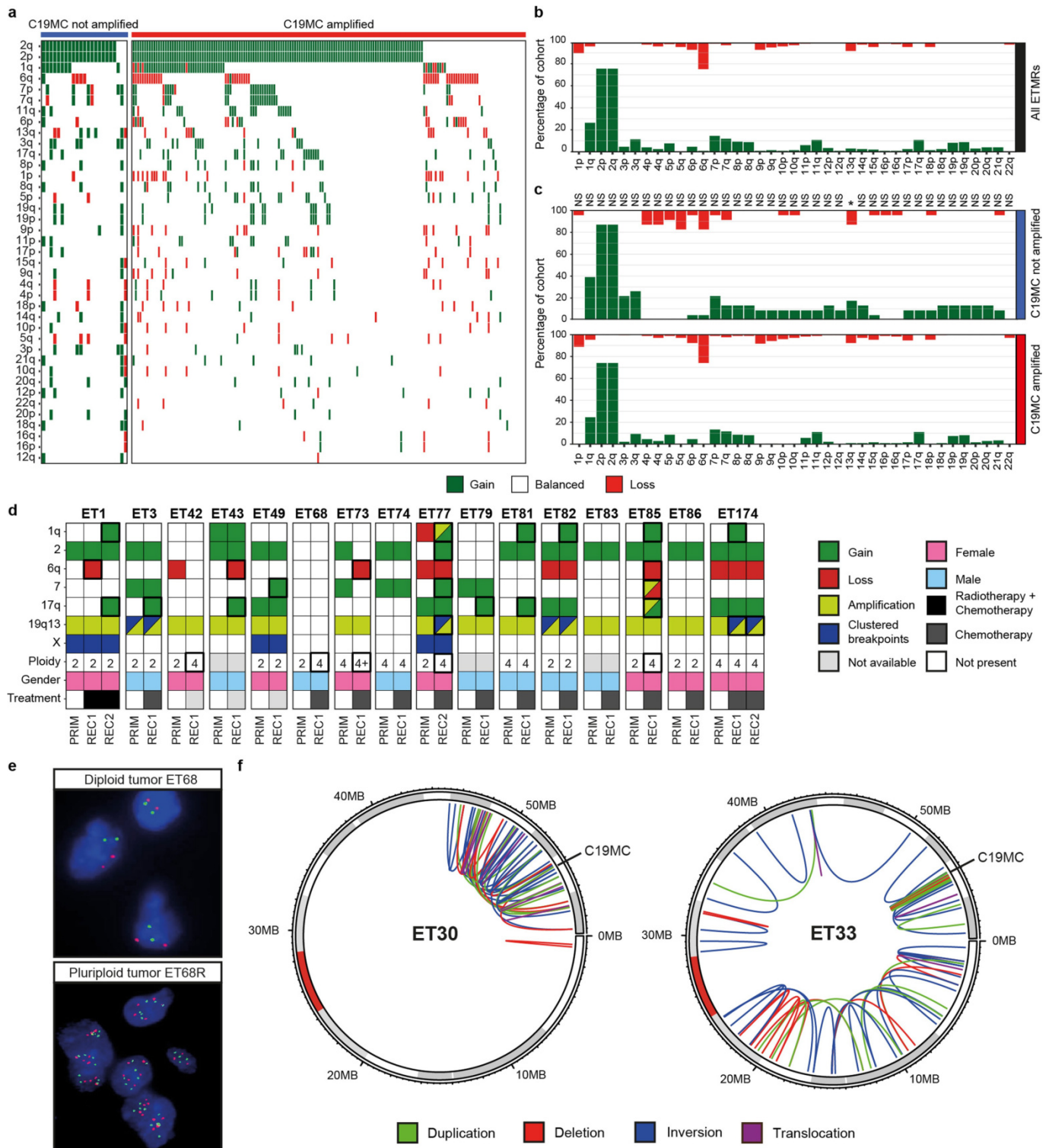
Extended Data Fig. 4 | ETMRs consist of at least two distinct subtypes of cells. **a**, Heat map showing z-score-normalized expression of 450 DNA repair genes and the corresponding pathways⁸ for 190 tumours of different entities including 28 ETMRs. Supervised clustering was used and samples were sorted by entity or C19MC amplification status. Entities include three ATRT subgroups, four MB subgroups, central nervous system ewing sarcoma family tumour with *CIC* alteration (CNS EFT-CIC), central nervous system neuroblastoma, with *FOXR2* activation (CNS-NB *FOXR2*), central nervous system high-grade neuroepithelial tumour with *MNI* alteration (HGNET-MN1), central nervous system high-grade neuroepithelial tumour with *BCOR* alteration (HGNET-BCOR), ETMRs with amplification of C19MC (red) and ETMRs without amplification of C19MC (blue). ETMR subsets were manually assessed based on DNA repair pathway expression. **b**, Debulking of mRNA expression using CIBERSort by using the median expression of single-cell RNA-sequencing data of the forebrain as gene signature¹⁰. The cumulative fraction of each cell type was calculated and samples were sorted according to the percentage of modelled neural stem cells. Samples were annotated based on

the subsets derived from **a**. **c**, Box plots showing expression of stem cell markers (*HMGA2*, *LIN28A*), astrocyte markers (*AQP4*, *GFAP*) and genes involved in the DNA damage response (*WEE1*, *CHEK2*) in ETMRs with high DNA repair expression ($n = 18$) and low DNA repair expression ($n = 10$). P values were calculated using a two-sided Mann-Whitney U -test; *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$; NS, not significant. Boxes show the median, first and third quartile, and whiskers extend to 1.5× the interquartile range. **d**, Distribution of histology annotation of 18 ETMRs for which these data were available divided into two subsets. The number of EBL phenotypes was significantly enriched in the high DNA repair expression group using a two-sided Fisher's exact test ($P = 3.7 \times 10^{-2}$). **e**, t -SNE clustering based on methylation profiles of a microdissected ETMR (ET174) (split in bulk, rosettes and neuropil) and 192 other ETMRs. **f**, Expression of *LIN28A* and *AQP4* in rosette tissue and neuropil tissue of the same tumour. **g**, Copy-number profiles of microdissected neuropil and rosettes from the same tumour. **h**, Fold change in expression of six markers in two matched recurrences normalized to the primary tumour.



Extended Data Fig. 5 | Recurrent events in ETMRs without C19MC amplification. **a**, Schematic representation of the translocation and amplification of a region on chromosome 11 with the host gene of the *miR-17-92* miRNA cluster (also known as *MIR17HG*) shown in red on chromosome 13. Regions were reconstructed using mate pair sequencing. The actual amplified region is circular denoted by arrows on each end. **b**, Copy-number profile of a tumour containing the *miR-17-92* cluster translocation and amplification. Copy numbers were derived from methylation array data with each dot representing a probe. Inset shows validation of both the chromosome 11 (*YAP1*; green) and chromosome 13 (*MIR17HG*; red) amplifications using FISH. **c**, Quantification of

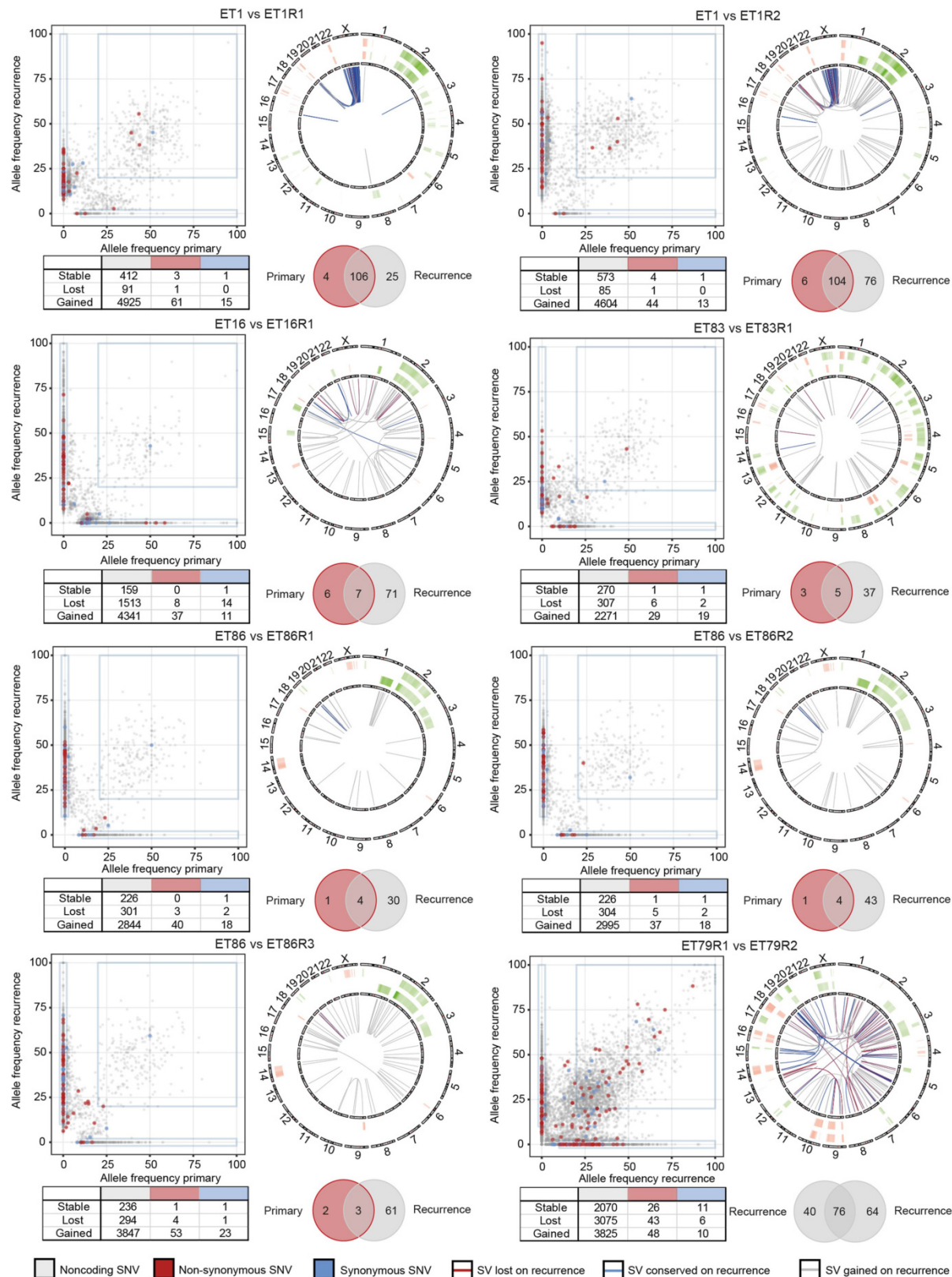
mature miRNAs in the *miR-17-92* miRNA cluster ($n = 20$) confirms that the ETMR (blue) with the chromosome 11 and chromosome 13 amplification and translocation has higher expression of *miR-17-92* cluster miRNAs. Each bar represents one tumour corresponding to the given entity. P values were calculated using a one-sided Mann-Whitney U -test; $*P < 0.05$. **d**, Example of a copy-number profile of a case showing clustered rearrangements around C19MC. This tumour did not have a C19MC amplification or *DICER1* mutations. **e**, Copy-number profile of an ETMR without C19MC amplification or *DICER1* mutation showing an overall instable genome with many regions containing clustered breakpoints.



Extended Data Fig. 6 | ETMRs recurrently show genomic instability.

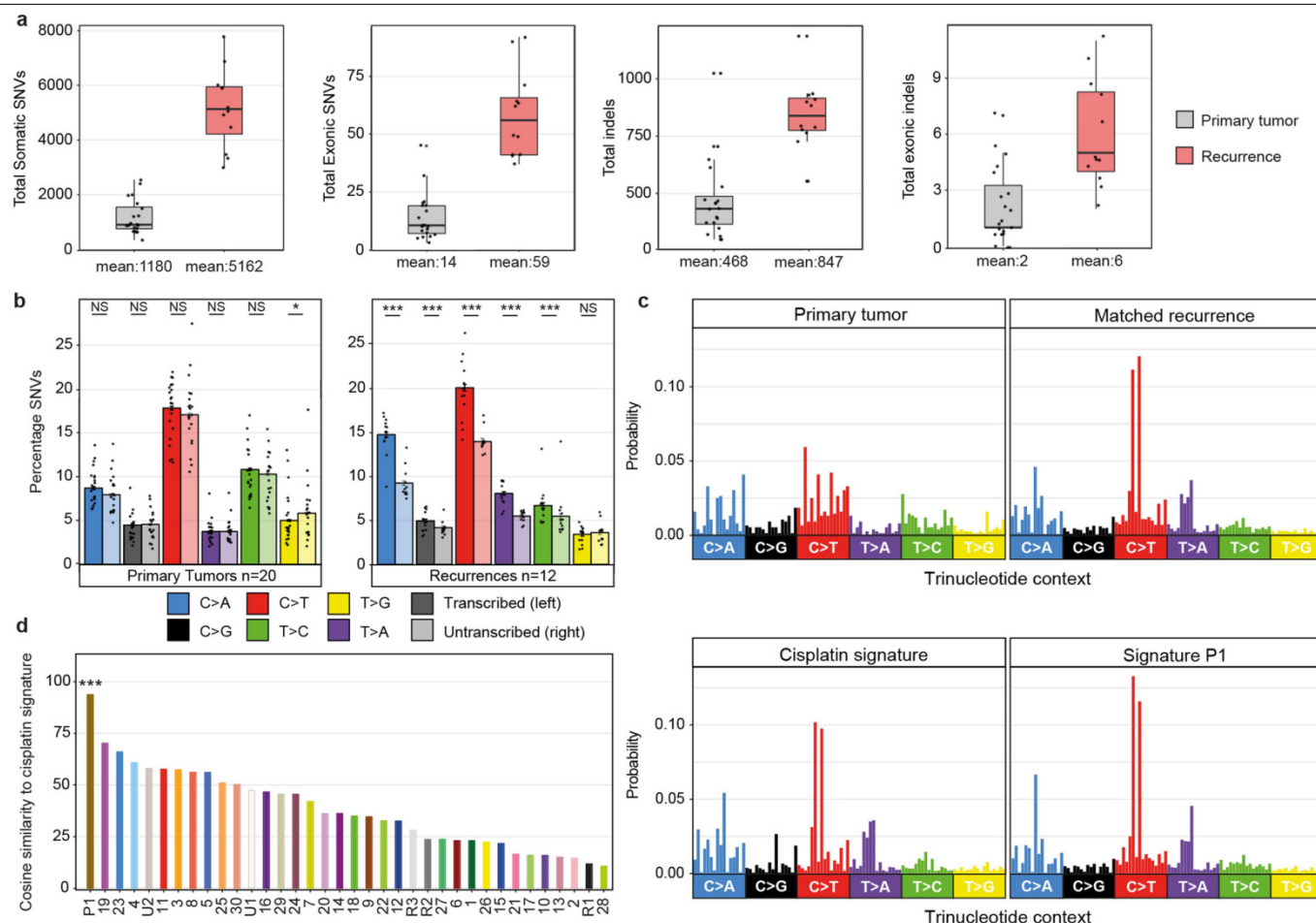
a, OncoPrint showing the co-occurrence of all CNAs separated by C19MC amplification status. **b**, Overview of copy-number profiles of all ETMRs ($n=193$). Bars (gain, balanced and loss) add up to 100% for each chromosome arm. **c**, Overview of copy-number profiles of all ETMRs with ($n=170$) or without ($n=23$) C19MC amplification. P values were calculated using two-sided Fisher's exact tests and adjusted for multiple testing (Benjamini-Hochberg correction); *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$. **d**, Overview of CNAs in matched primary tumour and recurrence pairs for the most variable CNAs.

Events (copy-number changes, clustered breakpoints or increases in ploidy) that were gained upon recurrence have a thicker outline. Percentages denote the percentage of matched samples acquiring a CNA or genome duplication. **e**, Example of a tumour for which polyploidy was validated using FISH ($n=28$ tested samples), the chromosome 9 and 11 centromeres were used as probes. **f**, Examples showing clustered breakpoints on chromosome 19. Chromosome 19 is shown as a circular representation, translocations to other chromosomes were annotated as single positions. All SVs were detected using mate pair sequencing.



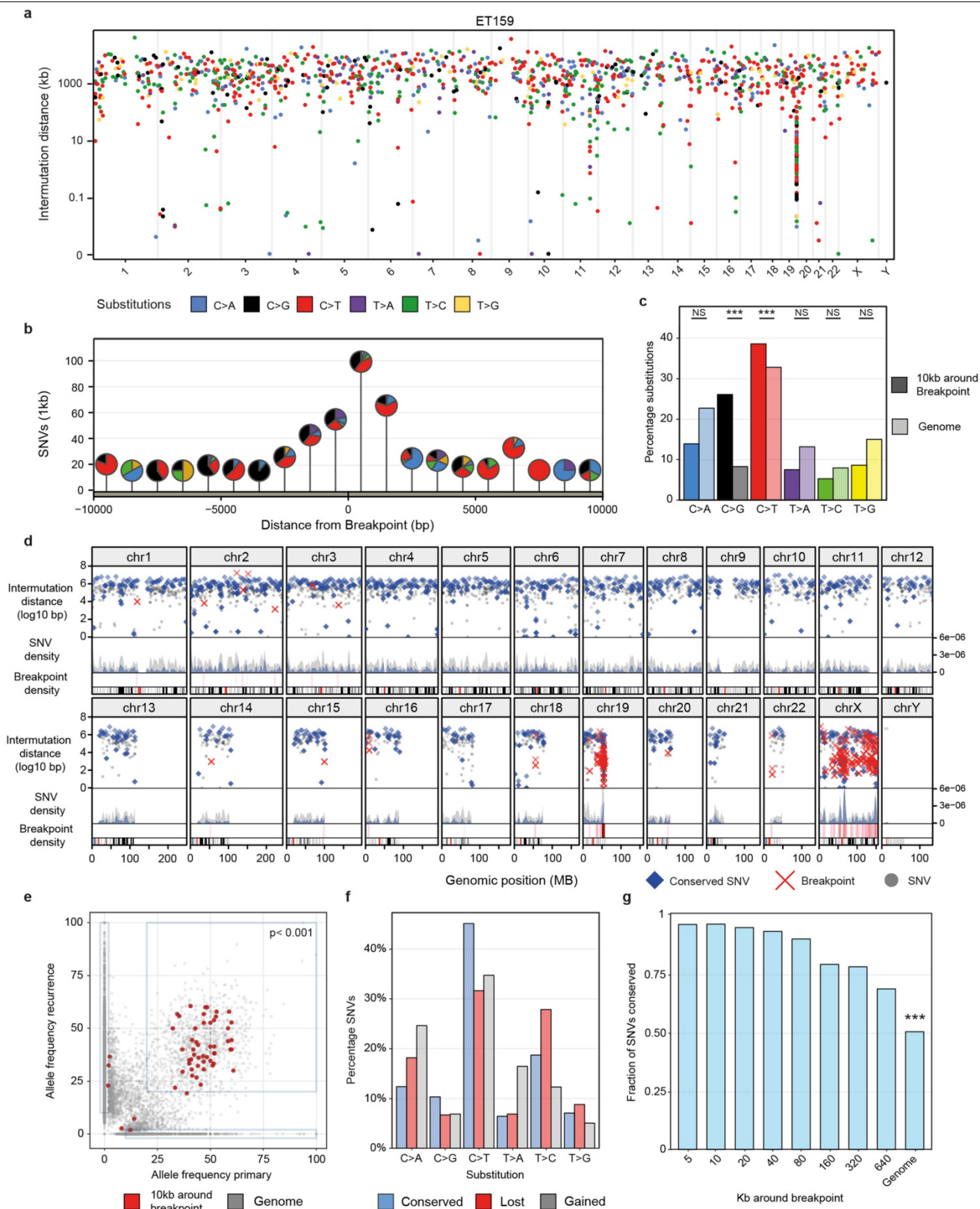
Extended Data Fig. 7 | Conservation of events for individual patients. Summary of events occurring in seven matched primary tumours compared to recurrences (first, second or third relapse) and two matched relapses. For every sample conservation of SNVs is given as a graph with the allele frequencies (AF) of the primary tumour on the x axis and the recurrence on the y axis. In the last panel, two matched recurrences are shown with a recurrence on each axis. Boxes show events that are lost, conserved or gained. Each comparison has a table showing the total number of events in each quadrant

(lost, primary AF > 10% and recurrence AF < 2%; stable, primary AF > 20% and recurrence AF > 20%; and gained, primary AF < 2% and recurrence AF > 10%). Conservation of SVs is given as a circular representation of the genome with the CNAs from the primary tumour in the outer rim and the recurrence in the inner rim. SVs were coloured by detection in either only the primary tumour (red), only in the relapse (grey) or in both (blue). Each combination also has a Venn diagram showing the total number of SVs that were detected in the primary tumour, the recurrence or both.



Extended Data Fig. 8 | Mutations in primary tumours and relapses. a, Box plots showing the total number of SNVs or indels in primary tumours ($n=20$) compared to relapses ($n=12$). Boxes show the median, first and third quartile, and whiskers extend to $1.5 \times$ the interquartile range. We detected, on average, 1,180 SNVs (range, 339–2,544) and 468 indels (range, 299–1,026) in primary tumours and 5,162 SNVs (range, 2,992–7,773) and 847 indels (range, 554–1,187) in relapsed tumours throughout the genome. In coding regions, there were on average 14 non-synonymous SNVs (range, 3–45) and 2 indels (range, 0–7) in primary tumours and 59 non-synonymous SNVs (range, 37–92) and 6 indels (range, 2–11) in relapsed tumours. **b**, Percentage of substitutions of either the combined primary tumours ($n=20$) or combined relapses ($n=12$) divided by

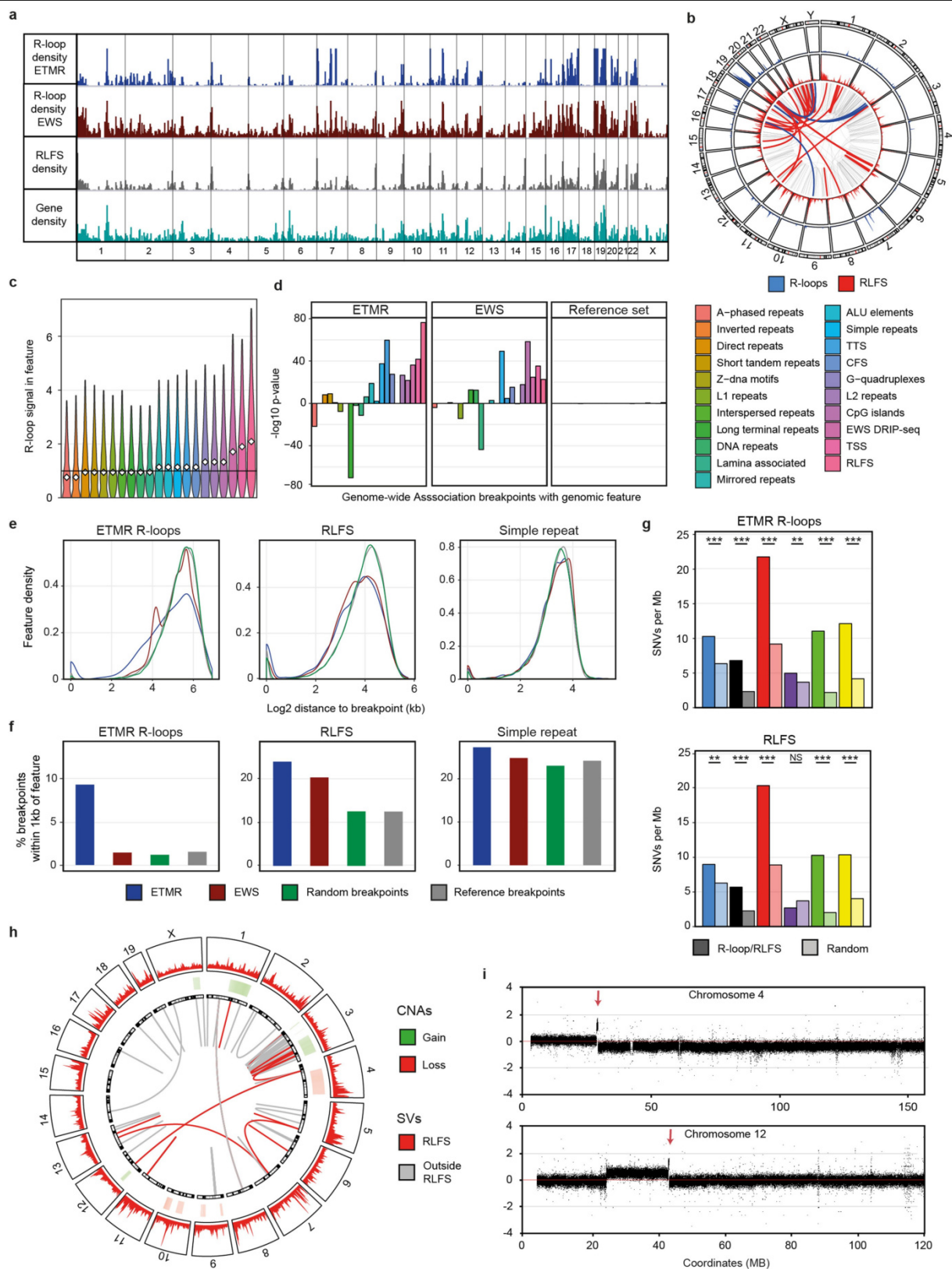
substitution type and affected strand for SNVs residing in transcribed regions. Transcriptional asymmetry is defined as the difference between the amount of SNVs on the transcribed strand versus the untranscribed strand for each substitution type. Data are mean \pm s.e.m., P values were calculated using two-sided Poisson tests; *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$. **c**, Substitution-type probability based on the 96 different trinucleotide contexts for a matched primary relapsed pair shown in **d** compared to a cisplatin signature¹⁶ and new paediatric cancer signature (P1)¹³. **d**, Cosine similarity between the cisplatin signature and other signatures ($n=36$). P values were calculated using pairwise Pearson correlation applied to the similarity matrix; *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | ETMRs have dense and strongly conserved C>T and C>G mutations around breakpoints. **a**, Rainfall plot showing an example of kataegis around C19MC. Every point represents a somatic SNV coloured by substitution type, the x axis represents the position in the genome and the position on the y axis represents the density of SNVs. **b**, Lollipop plot showing SNVs per 1 kb in a region of 10,000 bp surrounding breakpoints for all ETMRs. Pins represent the percentage of substitution types of all SNVs within 1 kb, while the height of the lollipops represents the substitutions per kb. **c**, Percentages of substitution types in regions 10 kb around breakpoints (left, $n = 543$ SNVs) and the rest of the genome (right, $n = 84991$ SNVs). P values were calculated using a one-sided Fisher's exact test; *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$. **d**, Combined mutation density of four primary tumours coloured by conservation in the matched recurrence (blue is conserved, grey is not conserved) as shown by a rainfall plot (top), a density distribution (middle) and the breakpoint density (bottom). **e**, Allele frequencies of all primary (x axis)

versus relapse (y axis) tumours. Boxes show conservation (lost, primary AF > 10% and recurrence AF < 2%; conserved, primary AF > 20% and recurrence AF > 20%; and gained, primary AF < 2% and recurrence AF > 10%) ($n = 2,100$ SNVs with allele frequency over 20% in the primary tumour). P value was calculated using a two-sided χ^2 test. **f**, Percentage of substitution types for SNVs in each quadrant (lost, primary AF > 10% and recurrence AF < 2%; conserved, primary AF > 20% and recurrence AF > 20%; and gained primary AF < 2% and recurrence AF > 10%). **g**, Ratio of conserved SNVs compared with not conserved SNVs in regions around breakpoints with increasing sizes. Conservation is defined as SNVs with an allele frequency over 20% in the primary tumour and an allele frequency over 20% in the recurrence, SNVs with an allele frequency lower than 20% in the recurrence but higher than 20% in the primary tumour were defined as not conserved. P value between 10 kb around breakpoints and the rest of the genome using a two-sided χ^2 test ($n = 2,100$, $P = 5.4 \times 10^{-11}$).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Context of R-loops and DNA damage in ETMRs and after *Dicer1* knockout. **a**, Genome-wide density of R-loops in ETMRs, R-loops in Ewing sarcoma (EWS), RLFS and gene density. **b**, Representation of SVs genome-wide and their breakpoint context. Outer layers show the density of DRIP peaks (blue) or RLFS (red). The inner part shows all SVs from ETMRs sequenced using WGS, depicting SVs that fall in DRIP-seq peaks (blue) or RLFS (red). **c**, R-loop signal detected in genomic regions sorted by R-loop signal (including elements from non-B-DB⁶² and repeatmasker). R-loop signal was determined for 10,000 randomly selected elements for every type of genomic feature ($n = 21$). Violin plots depict kernel density estimates and represent the density distribution. **d**, Genome-wide association of breakpoints with genomic regions sorted by R-loop signal shown in **c**. Genome-wide associations were calculated as distance to nearest element compared to a set of 10,000 randomly generated breakpoints. Enrichments were calculated for Ewing sarcoma breakpoints⁶⁶ and breakpoints from other entities²² (reference set). P values were calculated using a two-sided Mann–Whitney U -test and adjusted

for multiple testing (Benjamini–Hochberg correction). **e**, Density of distances between genomic regions and breakpoints detected in ETMR, Ewing sarcoma, random breakpoints and reference breakpoints. **f**, Total percentage of breakpoints within 1 kb of genomic regions. **g**, Enrichment of SNVs ($n = 85,534$) in ETMR R-loops ($n = 16,002$ regions) and RLFS ($n = 85,534$ regions) compared to random regions of the same size. P values were calculated using a two-sided χ^2 test; *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$. **h**, Genome-wide distribution of mouse RLFS and breakpoints occurring in *Dicer1* knockout cells compared to wild-type. The outer rim shows the genome wide density of mouse RLFS, the inner rim the CNAs that were found between wild-type and knockout cells and the inner part shows the SVs that were detected between wild-type and knockout cells. Breakpoints falling within RLFS are highlighted in red. **i**, Copy-number profiles of an example of a translocation coupled to duplication in RLFS that were found in *Dicer1* knockout compared to *Dicer1* wild-type cells. Red arrows depict the location of the translocation and duplication.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Software used for data collection was provided by the manufacturer (bcl2fastq (v2.2), QC performed with fastQC (0.11.4) (Illumina) or GCOS (3000) for microarray expression data (Affymetrix), minfi (v1.14.0), limma (v.3.24.15) and is described in the methods section. Data software to collect data using confocal microscopy was available from zeiss (Zen suite LSM 800).

Data analysis

All software used is described in the methods section. Besides custom scripts public software that was applied includes: biobambam (version 0.0.148), bwa-mem-0.7.8-r455, Roddy (version 1.1.73), Samtools 0.1.17-r973, Platypus (0.8.1), Delly (version 0.7.5), R (3.4.3), MACS (2.0), MASS.0, R2 (Revision 9f522207), Graphpad Prism (v6.01), ImageJ (1.52e), Cibersort (v1.06), Picard (v1.125), Freebayes (v1.1.0), vt (v0.5), Eland (v2), Varscan (version 2.3.9), DESeq2 (1.18.1), NaviGO, DAVID (v6.8), Toppgene (7), deepTools (3.0.2), qmRLFS finder (v1.5), conumee (v1.3), MutationalPatterns (v1.4.3), Compusyn, Bowtie (v1.00), IGV (v2.3.97), Rtsne (v0.13), SVA (3.32.1), kallisto (v0.43.0), Sleuth (v0.30.0) and Annovar (2016-02-01).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data used in this study will be made available before publication, as stated in the manuscript, within the corresponding public repository. Methylation and expression array data is available under accession number GSE122038 in the Gene Expression Omnibus (GEO) and sequencing data is available under EGAS00001003256 at the European Genome-phenome Archive (EGA). Source data are available for Fig. 1a-c, Fig. 2c, Fig. 3b, c, Fig. 4d, g, Fig. 5, a, b, d, Ext. Data Fig.

1a, Ext. Data Fig. 2a-g, Ext. Data Fig. 4b, c, h, Ext. Data Fig. 5c, Ext. Data Fig. 6b, c, Ext. Data Fig. 8a-d, Ext Data Fig. 9b, c, e, f, g and Ext Data Fig 10 g. Any additional data/material used in this study will be made available upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined based on the amount of available material, no sample size calculation was performed. We argue that the sample size used here is sufficient, first of all because ETMR tumor material is sparse and therefore it is difficult to expand the cohort and second of all because our sample size is comparable to other large genomic studies.
Data exclusions	Data was excluded from analysis if QC measures were not met. These measures were taken to prevent erroneous assumptions about the data that were caused by technical artifacts. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was applied supplemented with additional criteria after mapping including testing for equal coverage, the number of unmapped reads, presence of potential sample swaps based on SNP fingerprinting, poor read quality, uncertainty about the diagnosis of the sample, tumor cell content.
Replication	All experiments have been performed in triplicate and only representative data is shown. Dot blot experiments have been performed at least five times to ensure reproducibility. All attempts at reproducibility were successful
Randomization	No randomization was applied. All samples were included based on diagnosis as is stated in the methods section.
Blinding	Patient identity was blinded through anonymisation, however patient age, tumor location, histology, patient gender and treatment outcome were made available.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Anti-DNA-RNA hybrid (clone S9.6) antibody (Merck, MABE1095), yH2AX (Abcam, ab11174) Dilution: 1:200 lot #3011493
Validation	Antibodies were validated by the manufacturer as stated on the manufacturers website. Specificity of S9.6. was validated using both human and mouse material treated with RNASEH1 as an additional negative control.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The BT183 cell line was kindly provided by J.A.Chan. Dicer WT/KO cell lines were purchased from from ATCC (CRL-3220, CRL-3221)
Authentication	Cell lines were authenticated using similarity in variable SNP against the primary tumor in the case of the BT183 cell line and by using methylation clustering. Cell lines obtained from ATCC have been authenticated using STR profiling.

Mycoplasma contamination

All cells were found negative for mycoplasma infection

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified lines were used

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

All patient characteristics are available in supplementary table 1 if available. In general, the majority of patients was under the age of three at diagnosis, there was no gender bias (male:female=1:1), location of the tumors was evenly spread throughout the brain as is described in Ext. Data Figure 1b. Relapsed patients have been clearly annotated and were analyzed separately

Recruitment

Samples were collected from biopsies taken between 2000 and 2018 at contributing centers of ETMR patients that were treated at the respective centers. All patients included in the study were enrolled under informed consent according to the ICGC (www.icgc.org) guidelines or INFORM (www.dkfz.de/en/inform/) guidelines. No bias caused by recruitment is expected however possible under-diagnosis of patients lacking C19MC amplification (due to difficulty of diagnosis) cannot be excluded which could have lead to a lower proportion of patients lacking C19MC amplification.

Ethics oversight

Ethics oversight has been conducted by the ICGC and INFORM consortia

Note that full information on the approval of the study protocol must also be provided in the manuscript.

MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment

<https://doi.org/10.1038/s41586-019-1790-2>

Received: 10 March 2018

Accepted: 9 October 2019

Published online: 27 November 2019

Vincenzo Calvanese^{1,2*}, Andrew T. Nguyen¹, Timothy J. Bolan¹, Anastasia Vavilina¹, Trent Su³, Lydia K. Lee⁴, Yanling Wang¹, Fides D. Lay¹, Mattias Magnusson^{1,2}, Gay M. Crooks^{2,5,6}, Siavash K. Kurdistanj^{2,3,6,7} & Hanna K. A. Mikkola^{1,2,6,7*}

Limited knowledge of the mechanisms that govern the self-renewal of human haematopoietic stem cells (HSCs), and why this fails in culture, have impeded the expansion of HSCs for transplantation¹. Here we identify MLLT3 (also known as AF9) as a crucial regulator of HSCs that is highly enriched in human fetal, neonatal and adult HSCs, but downregulated in culture. Depletion of MLLT3 prevented the maintenance of transplantable human haematopoietic stem or progenitor cells (HSPCs) in culture, whereas stabilizing MLLT3 expression in culture enabled more than 12-fold expansion of transplantable HSCs that provided balanced multilineage reconstitution in primary and secondary mouse recipients. Similar to endogenous MLLT3, overexpressed MLLT3 localized to active promoters in HSPCs, sustained levels of H3K79me2 and protected the HSC transcriptional program in culture. MLLT3 thus acts as HSC maintenance factor that links histone reader and modifying activities to modulate HSC gene expression, and may provide a promising approach to expand HSCs for transplantation.

HSCs can self-renew throughout their lifetime while replenishing all blood lineages, making HSC transplantation a life-saving treatment for many blood diseases. However, a lack of HLA-matched bone marrow donors and a low yield of HSCs in cord blood limit the number of patients that can be treated¹. A better understanding of HSC self-renewal is required to expand human HSCs in culture or to generate them from pluripotent stem cells.

HSCs develop during embryogenesis from haemogenic endothelium in large arteries and expand in the fetal liver before colonizing the bone marrow². Although many factors that drive the specification of haemogenic endothelium and HSCs have been identified, we know less about those that maintain HSC self-renewal. Here we identify MLLT3 as a crucial regulator of human HSC maintenance, and show that restoring MLLT3 levels in cultured human HSCs protects stemness and enables the ex vivo expansion of transplantable HSCs.

MLLT3 is enriched and required in human HSCs

To define the molecular machinery that governs human HSC self-renewal and determine why it fails in culture, we compared the transcriptomes of highly self-renewing HSPCs from human fetal liver to their immediate progeny³ and to dysfunctional, cultured HSPCs, derived from fetal liver or embryonic stem cells^{4,5}. From the 12 nuclear regulators correlating with self-renewal, MLLT3 was selected for further study (Fig. 1a, Extended Data Fig. 1a, b). MLLT3 is a component of the superelongation complex⁶ and co-operates with DOT1L, which di/trimethylates H3K79 to promote transcription^{7–9}. MLLT3 localizes

to active transcription start sites (TSSs) through the YEATS domain, which recognizes active histone marks such as H3K9 acetylation and crotonylation^{8,10}. A truncated MLLT3 that lacks the YEATS domain forms a leukaemic fusion protein with the N terminus of MLL1, which misdirects MLLT3-interacting complexes to induce aberrant gene transcription^{11–14}. MLLT3 also regulates erythroid or megakaryocytic progenitors¹⁵ and was identified as a definitive HSC hub gene during mouse development¹⁶.

MLLT3 expression was enriched in undifferentiated human HSPCs in fetal liver, cord blood and bone marrow (Extended Data Fig. 1c). RNA-sequencing (RNA-seq) analysis of developmental haematopoietic tissues showed that *MLLT3* was upregulated in fetal liver¹⁷ (Extended Data Fig. 1d), whereas genes specific to the development of haemogenic endothelium and HSCs such as *TALI* (also known as *SCL*), *RUNX1*, *SOX17* and *HOXA* genes were already highly expressed in 5 week aorta-gonad-mesonephros, which suggests that MLLT3 is involved in HSC maturation and maintenance.

To determine whether human HSCs require MLLT3, two validated *MLLT3* short hairpin RNAs (shRNAs) (Extended Data Fig. 1e, f) were tested in an HSPC expansion culture system using the OP9M2 stromal stem-cell line⁴. Both of the shRNAs resulted in premature depletion of fetal liver HSPCs (FL-HSPCs) in vitro (Fig. 1b, c, Extended Data Fig. 1g–j). When FL-HSPCs transduced with MLLT3-knockdown (KD) or control vector were transplanted into immunodeficient NSG (NOD-SCID *IL2rg*-null) mice, only the control cells showed multilineage (myelo/lymphoid) human haematopoietic reconstitution (Fig. 1d, e, Extended Data Fig. 1k, Supplementary Table 1), which indicates an important regulatory function for MLLT3.

¹Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA. ²Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA, USA. ³Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA, USA. ⁴Department of Obstetrics and Gynecology, University of California Los Angeles, Los Angeles, CA, USA. ⁵Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ⁶Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA, USA. ⁷Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA. *e-mail: vincalv@gmail.com; hmikkola@mcdb.ucla.edu

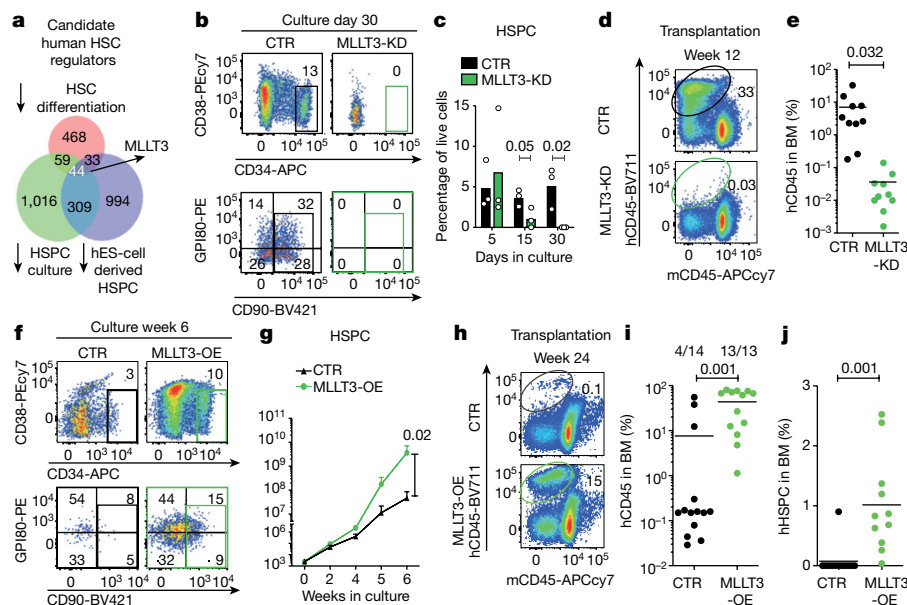


Fig. 1 | MLLT3 regulates human HSPC expansion. **a**, Venn diagram of microarray gene expression data, identifying genes enriched in self-renewing human FL-HSPCs. Number of genes downregulated after differentiation (pink) of fetal liver CD34⁺CD38^{-/-}CD90⁺GPI80⁺ HSCs to CD34⁺CD38^{-/-}CD90⁺GPI80⁺ progenitors³; number of genes downregulated in FL-HSPCs during 5-week culture on OP9M2 stroma (green)⁴; and number of genes suppressed in human embryonic stem (ES)-cell-derived HSPCs (purple)⁵ are shown. **b**, FACS analysis 30 days after transduction of CD34⁺CD38^{-/-}CD90⁺ HSPCs with MLLT3 shRNA (MLLT3-KD) or empty vector control (CTR) (representative of three plots). **c**, Quantification of cells as in **b** after 5, 15 and 30 days in culture ($n = 3$). **d**, FACS analysis of bone marrow from NSG mice 12 weeks after transplantation of FL-HSPCs transduced with MLLT3-KD or empty vector control (representative of 10 mice). **e**, Quantification of human (h) CD45⁺ cells in bone marrow (BM) from

NSG mice treated as in **d** ($n = 10$ mice, two independent experiments). **f**, FACS analysis of CD34⁺CD38^{-/-}CD90⁺ FL-HSPCs transduced with control or MLLT3-OE lentiviral vector (representative of six experiments). **g**, Expansion of HSPCs as in **f** ($n = 6$ independent experiments). **h–j**, FACS analysis showing human haematopoietic reconstitution (hCD45 expression) (**h**) and quantification of total hCD45 cells (**i**) or human HSPCs (hCD45⁺CD34⁺CD38^{-/-}) (**j**) in bone marrow from NSG mice. In **i** and **j**, the number of mice showing multi-lineage reconstitution versus the number of total transplanted mice is shown ($n = 13$ or 14 mice, 4 independent experiments). APCcy7, PEcy7, BV421 and BV711 denote fluorochrome dyes. Data in **c** denote mean values; data in **e**, **i** and **j** denote mean and individual values; data in **g** are mean \pm s.e.m. All P values determined by two-sided t -test.

Sustaining MLLT3 levels improves HSPC culture

As studies had shown expansion of multipotent human HSPCs on OP9M2 stroma without measurable expansion of transplantable HSCs⁴, we asked whether maintaining MLLT3 expression in cultured HSPCs improves their function. Restoring physiological MLLT3 levels using an overexpression lentiviral vector (MLLT3-OE) significantly enhanced the expansion of CD34⁺CD38^{-/-}CD90⁺ FL-HSPCs on OP9M2 stroma (78-fold greater than controls at 6 weeks) (Fig. 1f, g, Extended Data Fig. 2a, b). Targeting MLLT3-OE in the most undifferentiated CD34⁺CD38^{-/-}CD90⁺GPI80⁺ FL-HSPCs³ recapitulated the expansion phenotype, whereas MLLT3-OE in CD34⁺CD38^{-/-}CD90⁺GPI80⁺ progenitors did not confer the HSC immunophenotype or expand them (Extended Data Fig. 2c–e). Withdrawing OP9M2 stroma depleted undifferentiated MLLT3-OE HSPCs, demonstrating a dependence on a HSC-supportive microenvironment (Extended Data Fig. 2f–h). MLLT3-OE also enhanced FL-HSPC expansion when cultured in clinically suitable conditions using serum-free expansion medium (SFEM) and the small molecules SR1¹⁸ and UM171¹⁹, or on OP9M2 stroma with both SR1 and UM171 (Extended Data Fig. 2i, j), showing the beneficial effects on HSPC expansion on all HSC-supportive conditions tested.

BrdU incorporation assays did not show enhanced proliferation of MLLT3-OE HSPCs in culture (Extended Data Fig. 3a, b). Staining of annexinV and 7-aminoactinomycin D (7AAD) showed greater viability of MLLT3-OE FL-HSPCs than empty-vector-transduced FL-HSPCs, but reduced cell viability compared with uncultured FL-HSPCs (Extended Data Fig. 3c). Differentiation of MLLT3-OE HSPCs after a 4-week expansion showed comparable monocytic, granulocytic, erythroid, megakaryocytic, T and B lymphoid differentiation potential to non-expanded

FL-HSPCs (Extended Data Fig. 3d, e). Hence, sustaining MLLT3 expression in cultured HSPCs maintains their identity and viability without causing excessive proliferation or resistance to apoptosis, or blocking differentiation.

MLLT3 enhances transplantation of cultured HSPCs

To assess whether maintaining MLLT3 levels in cultured HSPCs improves the reconstitution potential in NSG mice, CD34⁺GFP⁺ fetal liver cells sorted at day 5 and transduced with MLLT3-OE or control vector were expanded in SFEM containing SR1 and UM171 and transplanted on day 15 (Extended Data Fig. 4a). Mice transplanted with MLLT3-OE cells showed more frequent long-term (24 week) human multilineage engraftment in the bone marrow, and higher levels of engraftment than mice transplanted with control HSPCs (Fig. 1h, i, Extended Data Fig. 4b, Supplementary Table 2a). Most mice transplanted with MLLT3-OE cells, but not with control cells, also contained human HSPCs in the bone marrow and generated myeloid and lymphoid cells in the peripheral blood and spleen (Fig. 1j, Extended Data Fig. 4c–e). Increased haematopoietic reconstitution with MLLT3-OE HSPCs was not explained by altered proliferation of MLLT3-OE HSPCs or differentiated cells in recipient mice or in culture before transplantation (Extended Data Fig. 4f–j).

To ascertain the role of extended culture in MLLT3-mediated enhanced in vivo reconstitution, the engraftment levels were compared between CD34⁺GFP⁺ cells sorted 5 days after transduction and their progeny sorted at 15 days. Culture with MLLT3-OE, but not control vector, significantly increased total human haematopoietic reconstitution and HSPC reconstitution (Extended Data Fig. 4k–m). When bone marrow from primary mice was transplanted into secondary mouse

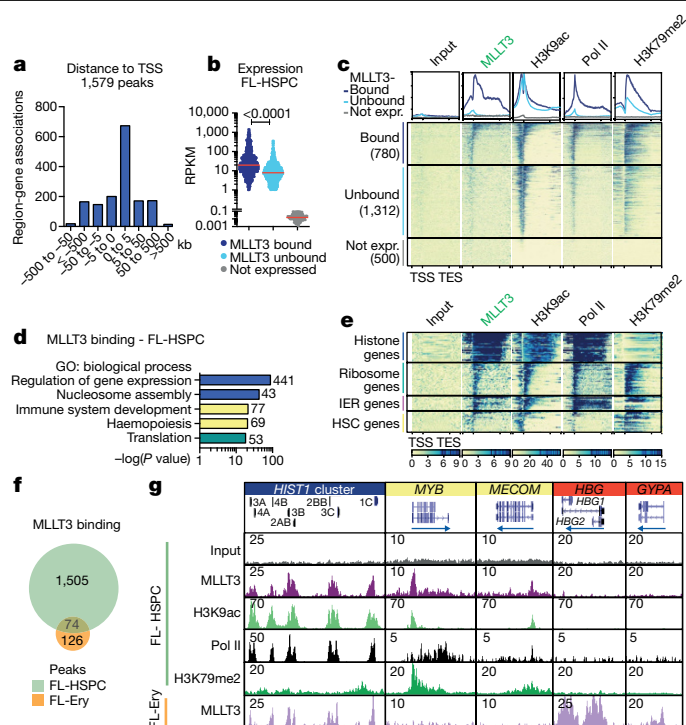


Fig. 2 | MLLT3 binds to the TSSs of active genes in a cell-type-specific manner. **a**, Distribution of MLLT3 peaks in FL-HSPCs. **b**, Genes expressed in FL-HSPCs (RPKM > 1, $n = 3$) were divided into MLLT3-bound genes (780, peak < 5 kb from the TSS) and MLLT3-unbound genes (1,312, no detectable peak), and compared to 500 randomly selected non-expressed genes (RPKM < 0.1). Red lines denote median values. P value determined by two-sided t -test. **c**, Average profile and heat map for MLLT3, H3K9ac, RNA Pol II and H3K79me2 ChIP-seq in FL-HSPCs. Metagene plot ± 2 kb is shown. **d**, GO analysis of MLLT3-bound genes in FL-HSPCs with gene numbers for each category. **e**, Heat maps for selected MLLT3-bound gene groups. IER, immediate early response. **f**, Venn diagram comparing MLLT3 peaks in FL-HSPCs and fetal liver erythroblasts (FL-Ery). **g**, UCSC genome browser tracks showing ChIP-seq of MLLT3 and epigenetic marks in representative MLLT3-bound genes in FL-HSPCs and FL-erythroblasts. In **a–g**, $n = 3$ MLLT3 in FL-HSPCs, $n = 2$ others.

recipients, only MLLT3-OE cells showed multilineage reconstitution after 18 weeks (Extended Data Fig. 4n, o). These data indicate that maintaining MLLT3 expression in FL-HSPCs during culture enhances human multilineage haematopoietic reconstitution in primary and secondary recipients.

MLLT3 binds to TSSs of active genes in human HSPCs

To understand how MLLT3 regulates human HSC stemness, we assessed the MLLT3 chromatin-binding pattern in FL-HSPCs. Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) showed MLLT3 binding at 1,579 sites, with strongest enrichment around TSSs and within 5 kb downstream (Fig. 2a). MLLT3 peaks associated with 889 genes, 96.4% of which were expressed in FL-HSPCs, with reads per kilobase of transcript per million mapped reads (RPKM) values greater than one (Supplementary Table 3). The k -means clustering analysis showed co-localization of MLLT3 peaks with marks of active TSSs (assay for transposase-accessible chromatin using sequencing (ATAC-seq) peaks, H3K4me3, H3K9ac, H3K9cr, H3K27ac and RNA polymerase (Pol II)) (Extended Data Fig. 5a). The gene body histone mark H3K79me2 partially overlapped with MLLT3, whereas H3K36me3 was not enriched. There was minimal overlap with the enhancer mark H3K4me1 or the repressive marks H3K27me3 and

H3K9me3. Thus, MLLT3 predominantly localizes to active promoters in HSPCs.

Comparing MLLT3-bound genes to other expressed genes showed that, although both contained active epigenetic marks at TSSs, MLLT3-bound genes featured higher median expression, higher H3K79me2 enrichment and higher RNA Pol II occupancy (Fig. 2b, c, Extended Data Fig. 5b). Gene Ontology (GO) analysis of MLLT3-bound genes in FL-HSPCs revealed enrichment of biological processes involved in regulation of gene expression and nucleosome assembly (for example, histone genes), immune system development and haemopoiesis (for example, HSC transcription factors RUNX1, MYB, MECOM and HOXA9) and translation (for example, ribosomal proteins) (Fig. 2d, Supplementary Table 3a–c). Analysis of epigenetic marks in distinct MLLT3-bound gene groups revealed differential enrichment for H3K79me2 and Pol II: the histone genes and immediate early response genes (such as *JUN* and *FOS*) showed high Pol II occupancy but low H3K79me2, whereas HSC genes and ribosomal protein genes showed high enrichment for H3K79me2 (Fig. 2e, Extended Data Fig. 5c). These data suggest that MLLT3 may regulate distinct target genes in HSPCs by influencing H3K79me2 and/or Pol II activity.

Analysis of MLLT3 binding in erythroblasts from fetal liver (Fig. 2f, Supplementary Table 3d) revealed cell-type specificity: 200 MLLT3 peaks were identified around TSSs, partially overlapping with FL-HSPC peaks (Extended Data Fig. 6a–c). Common GO categories included nucleosome assembly, whereas erythroid-specific categories included oxygen transport and haem metabolic processes (Extended Data Fig. 6d). Genome browser tracks demonstrated the correlation of MLLT3 binding with epigenetic marks of active TSS in each cell type (Fig. 2g, Extended Data Fig. 6a), and enrichment of H3K79me2 in MLLT3-bound HSC transcription factor genes.

MLLT3 protects HSC gene expression in cultured HSPCs

We next asked how sustaining MLLT3 expression in cultured FL-HSPCs modulates their transcriptional program. ChIP-seq data showed similar distribution of MLLT3-bound peaks and genes in uncultured HSPCs and MLLT3-OE-vector transduced HSPCs after a 4-week culture (Extended Data Fig. 7a–c). RNA-seq showed that modest differences in MLLT3 expression (2.77-fold MLLT3-OE versus control vector in HSPCs after 4-week culture) resulted in significant differential expression of 541 upregulated and 717 downregulated genes (Fig. 3a, Supplementary Table 4). Genes that regulate translation and glycolysis, and several HSC transcription factors (*MECOM* (also known as *EVII*), *HLF*, *MYB* and *GFI1*) and HSC surface proteins (*c-KIT* (also known as *KIT*), *CXCR4*, *ROBO4*, *EMCN* and *PROM1* (also known as *CD133*)) were significantly upregulated in MLLT3-OE HSPCs (Fig. 3b). Programs related to immune response and apoptosis were suppressed (Fig. 3b). Although MLLT3 binding was equally distributed between upregulated and downregulated genes, MLLT3 binding was enriched in specific gene categories such as MLLT3-OE-upregulated HSC transcription factor genes, and downregulated nucleosome assembly (for example, histone) genes. Immune response and apoptosis genes suppressed in MLLT3-OE HSPCs showed minimal binding (Fig. 3b–d). Comparison of cultured and uncultured HSPCs suggested that MLLT3-OE may help to diminish culture-associated drift in gene expression, either directly (HSC factors) or indirectly (immune response genes)⁴ (Fig. 3d, e, Extended Data Fig. 7d). Functional assessment of 2 of the 12 candidate HSC factors identified in Extended Data Fig. 1a—*MECOM* and *HLF*—both bound and upregulated by MLLT3, validated them as important MLLT3 downstream effectors that sustain HSC stemness. Knockdown of either factor resulted in premature HSPC exhaustion and diminished the effects of MLLT3-OE on HSPC expansion (Extended Data Fig. 7e–h).

Given the strong association between MLLT3 binding and H3K79me2 deposition in genes that encode HSC regulators, we asked whether MLLT3-OE protects the expression of HSC genes in cultured HSPCs

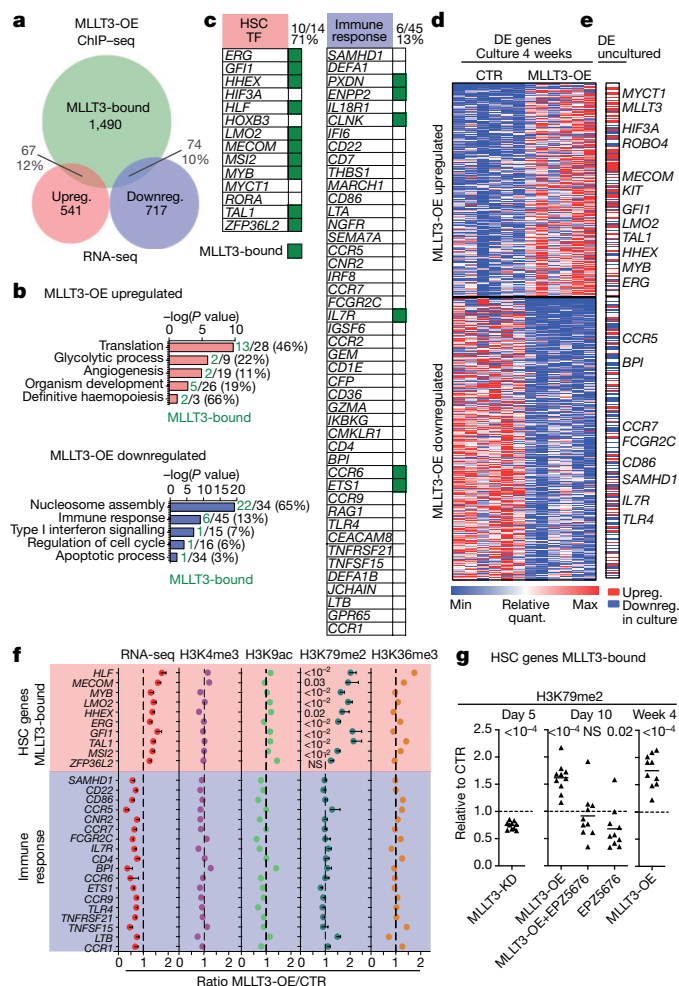


Fig. 3 | MLLT3 protects HSC stemness program through DOT1L and H3K79me2. **a**, Venn diagram showing the overlap between MLLT3-bound ($n = 6$ MLLT3 ChIPs combined; Extended Data Fig. 7b) and MLLT3-OE up- or downregulated genes. Total number and percentage of genes are shown. **b**, GO analysis of genes up- or downregulated in MLLT3-OE FL-HSPCs at 4 weeks. Numbers and percentage of MLLT3-bound and total genes are shown. **c**, Examples of gene groups up- or downregulated by MLLT3-OE with MLLT3-bound genes marked in green. Numbers and percentage of MLLT3-bound and total genes are shown. **d**, Heat map showing differentially expressed genes in HSPCs transduced with MLLT3-OE or control vector after 4-week expansion ($n = 6$ independent experiments). $P < 0.05$, Benjamini–Hochberg adjusted t -test. **e**, Differential expression (DE) of MLLT3-regulated genes between uncultured HSPCs ($n = 3$) and 4-week expanded FL-HSPCs ($n = 6$). Selected MLLT3-regulated genes that are similarly regulated in uncultured HSPCs are highlighted. MLLT3-bound genes are in bold. **f**, Quantification of RNA-seq and ChIP-seq signals in MLLT3-bound, upregulated HSC transcription factor genes, and unbound, downregulated immune response genes. RNA-seq represents fold change between expanded HSPCs transduced with MLLT3-OE or control vector. ChIP-seq shows ratio of MLLT3-OE and control signal normalized to non-MLLT3-bound housekeeping gene (*VCL*) ($n = 6$ RNA-seq; $n = 5$ H3K79me2 ChIP, $n = 2$ other ChIPs). Data are mean \pm s.e.m. P values determined by two-sided t -test. **g**, Ratio of H3K79me2 signal in MLLT3-bound upregulated HSC genes between MLLT3-KD and control vector after 5-day culture (data are mean and individual values, $n = 10$ genes, two experiments) (left), and between MLLT3-OE, MLLT3-OE plus EPZ5676 (DOT1L inhibitor), or CTR plus EPZ5676 and control vector after 10-day culture (middle). H3K79me2 signal is normalized to S2 cell chromatin spike-in. MLLT3-OE versus control from 4-week culture is also shown (right). P values in **g** determined by one-sample two-sided t -test. NS, not significant.

through DOT1L and H3K79me2. A significant MLLT3-dependent increase in H3K79me2 was observed in MLLT3-bound haematopoietic regulators but not in immune response genes that were indirectly downregulated by MLLT3-OE (Fig. 3f, Extended Data Fig. 8a). Other active marks tested (H3K9ac, H3K4me3 and H3K36me3) showed little change with MLLT3-OE. Dependence of H3K79me2 on MLLT3 levels was verified in MLLT3-KD HSPCs (Fig. 3g). Culture with the DOT1L inhibitor EPZ5676²⁰ reduced MLLT3-OE-associated increase in H3K79me2 in HSC genes (Fig. 3g). EPZ5676 decreased H3K79me2 in both MLLT3-bound and non-bound genes, whereas MLLT3-KD and MLLT3-OE only affected H3K79me2 in MLLT3-bound genes (Extended Data Fig. 8b). These data suggest that, although MLLT3 is not required for DOT1L activity and H3K79me2 deposition per se, MLLT3 cooperates with DOT1L in human HSCs to enhance H3K79me2 deposition in HSC regulatory genes and to maintain their activity during culture expansion.

MLLT3 enables ex vivo expansion of cord blood HSCs

The low number of HSCs in cord blood limits their use for transplantation, despite their better availability and more permissive HLA-matching than bone marrow. We therefore asked whether MLLT3 can be used to control the self-renewal of cord-blood HSCs (CB-HSCs). MLLT3-KD in cord-blood HSPCs (CB-HSPCs) severely impaired HSPC maintenance in co-culture experiments with OP9M2 stroma cells (Extended Data Fig. 9a, b). Conversely, MLLT3-OE improved CB-HSPC expansion both in co-culture with OP9M2 cells and in serum-free conditions with UM171 and SR1 (Fig. 4a, b, Extended Data Fig. 9c, d).

A limiting-dilution transplantation assay (Fig. 4a) was used to quantify the expansion of transplantable HSCs (repopulating units) during CB-HSC culture. Several doses of MLLT3-OE-transduced and control-transduced cells were transplanted either at day 5 after transduction or day 15 after culture and compared to uncultured CB-HSPCs. When high doses were transplanted, multilineage human haematopoietic reconstitution was observed in mice transplanted with uncultured cells or with HSPCs transduced with control vector or MLLT3-OE vector (Fig. 4c). Quantification of lineage differentiation in engrafted mice confirmed a comparable differentiation ability of MLLT3-OE and uncultured HSPCs (Fig. 4d), whereas HSPCs expanded by control vector showed myeloid bias. Moreover, mice engrafted with MLLT3-OE cells showed both increased HSPCs and total reconstitution compared with equal cell doses of control cells (Fig. 4e, f). Analysis of mice transplanted at limiting dilutions verified robust expansion of MLLT3-OE HSCs at 15 days of culture; 12.5-fold increase in repopulating units compared with uncultured HSPCs and a 6.8-fold increase compared with MLLT3-OE HSPCs transplanted at day 5 (Fig. 4f–h). Therefore, increased reconstitution ability was not caused merely by the expansion of MLLT3-OE cells in recipient mice. Although cells transduced with control vector showed some expansion (2.4-fold compared with uncultured HSPCs) at 15 days, the expansion of repopulating units with MLLT3-OE was 5.2-fold higher. These data corroborate the importance of maintaining MLLT3 levels during culture, both to achieve greater HSC expansion and preserve differentiation potential.

Blood and spleen showed solid human haematopoietic reconstitution by MLLT3-OE-expanded cells without evidence of lineage bias, whereas control-vector-expanded cells showed inconsistent, lower-level engraftment (Extended Data Fig. 10a–d). Secondary transplantation of bone marrow from primary mice resulted in human haematopoietic reconstitution with uncultured and MLLT3-OE cells, but not with control-vector-expanded cells (Extended Data Fig. 10e, f). Neither primary nor secondary mice transplanted with MLLT3-OE cells suffered increased mortality or expansion of immature populations (Extended Data Fig. 10g, h), unlike mice transplanted with haematopoietic cells expressing the oncogenic fusion protein MLL1–MLLT3 that rapidly developed leukaemia^{21,22}. Thus, sustaining MLLT3 expression in CB-HSPCs during culture enables a more than 12-fold expansion of

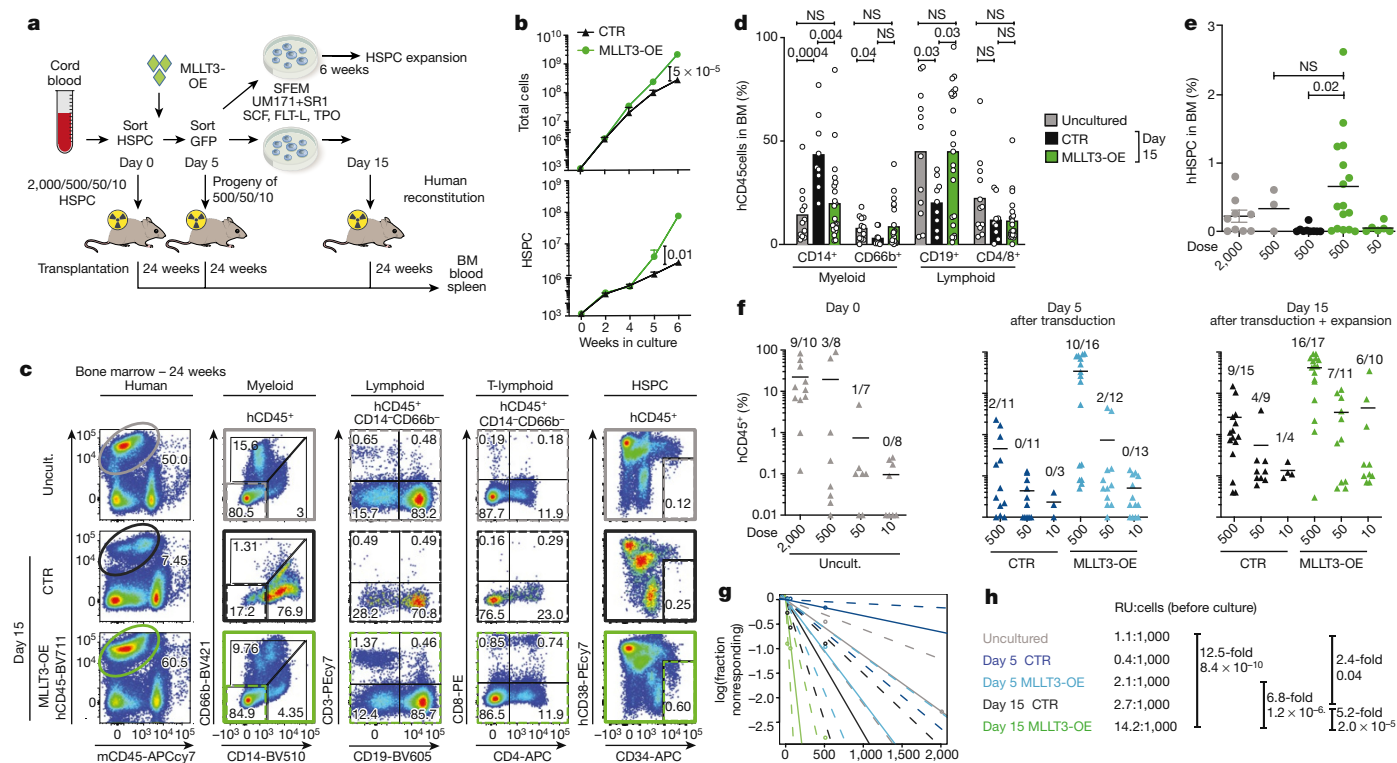


Fig. 4 | MLLT3-OE expands transplantable CB-HSCs in culture. **a**, Limiting dilution analysis for assessing MLLT3-OE effects on CB-HSPC expansion and in vivo reconstitution. **b**, Quantification of total cell and HSPC expansion in culture ($n = 3$). Data are mean values. P values determined by two-sided t -test. **c**, Representative FACS plots showing human haematopoietic reconstitution in mice transplanted with uncultured CB-HSPCs or CB-HSPCs transduced with control or MLLT3-OE vectors and transplanted at day 15 after culture and analysed 24 weeks after transplantation. FACS plots show total human haematopoietic reconstitution (CD45), multilineage differentiation (myeloid, B-lymphoid and T-lymphoid cells) and HSPCs in hCD45⁺ cells. **d**, Quantification of differentiated populations as a percentage of hCD45⁺ cells in the bone marrow of engrafted mice. **e**, Quantification of HSPCs as a percentage of live bone marrow cells in engrafted mice. **f**, Quantification of human

haematopoietic reconstitution in NSG mice at 24 weeks. Mice were transplanted in limited dilution doses with uncultured cells, or with transduced cells cultured for 5 or 15 days. The percentage of hCD45⁺ cells in bone marrow and the number of multilineage engrafted versus total transplanted mice is indicated. Data in **d–f** are mean and individual values from three independent experiments. P values determined by two-sided t -test. In **f**, n denotes engrafted mice/total transplanted mice for each condition. **g**, Calculation of reconstituting units from limiting dilution assay. Data in **h** are expressed as reconstituting units (RU) per 1,000 uncultured CB-HSPCs transplanted directly or transduced with MLLT3-OE or control vector and cultured for 5 or 15 days before transplantation. Expansion factors and P values between reconstituting units in uncultured HSPCs and MLLT3-OE or control vector-expanded cells were calculated using ELDA software²⁹ and are shown.

transplantable HSCs that maintain balanced multilineage haematopoiesis, demonstrating self-renewal of human HSCs in culture.

Discussion

Our discovery that restoring MLLT3 levels improves the expansion of engraftable, multipotent human HSCs when cultured with SR1 and UM171—small molecules that show promise for clinical HSC expansion^{19,23,24}—suggests that combining several strategies to support HSCs may be optimal for clinical translation. Notably, recent studies showed a marked increase in mouse HSC activity after replacing albumin in culture medium²⁵. As harnessing HSC self-renewal in culture becomes a reality, the potential for transformation and clonal haematopoiesis needs to be monitored carefully, independent of the expansion protocol used. Because every cell division in itself poses a risk for mutations, the goal should not necessarily be maximal HSC expansion, but safe expansion that preserves HSC integrity. Although no adverse effects were observed with the MLLT3-OE vector, the use of transient, non-integrating methods to maintain MLLT3 levels would be more appropriate for clinical use.

Our work identified MLLT3 as an ‘HSC maintenance factor’ that preserves—rather than confers—HSC stemness, by reinforcing programs established by other factors. Localization of MLLT3 to active TSSs of HSC genes in human HSCs in a cell-type-specific manner, using its

YEATS domain^{8,10}, provides a molecular basis for the accurate binding of MLLT3-OE in cultured HSPCs. This explains why full-length MLLT3 does not induce a differentiation block or ectopic activation of self-renewal, unlike the oncogenic MLL1–MLLT3 fusion protein, which rapidly converts haematopoietic progenitors to self-renewing leukaemia stem cells^{12,21,22}. Our work identifies MLLT3 as a central regulator of transcription factors that individually control HSC function^{26–28} and suggests that MLLT3–DOT1L-dependent regulation of H3K79me2 helps to maintain an active chromatin state in HSC regulatory genes during culture. With the ability to protect HSC stemness program as HSCs divide, without enhancing proliferation or imposing self-renewal program on progenitors, MLLT3 may help to improve HSC ex vivo expansion for clinical use.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1790-2>.

1. Ballen, K. K., Gluckman, E. & Broxmeyer, H. E. Umbilical cord blood transplantation: the first 25 years and beyond. *Blood* **122**, 491–498 (2013).

2. Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis. *Cell. Mol. Life Sci.* **73**, 1547–1567 (2016).
3. Prashad, S. L. et al. GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. *Cell Stem Cell* **16**, 80–87 (2014).
4. Magnusson, M. et al. Expansion on stromal cells preserves the undifferentiated state of human hematopoietic stem cells despite compromised reconstitution ability. *PLoS ONE* **8**, e53912 (2013).
5. Dou, D. R. et al. Medial *HOXA* genes demarcate haematopoietic stem cell fate during human development. *Nat. Cell Biol.* **18**, 595–606 (2016).
6. He, N. et al. Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. *Proc. Natl Acad. Sci. USA* **108**, E636–E645 (2011).
7. Steger, D. J. et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol. Cell Biol.* **28**, 2825–2839 (2008).
8. Li, Y. et al. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **159**, 558–571 (2014).
9. Bitoun, E., Oliver, P. L. & Davies, K. E. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum. Mol. Genet.* **16**, 92–106 (2007).
10. Li, Y. et al. Molecular coupling of histone crotonylation and active transcription by AF9 YEATS domain. *Mol. Cell* **62**, 181–193 (2016).
11. Schoch, C. et al. AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogenetically analyzed AML cases. *Blood* **102**, 2395–2402 (2003).
12. Krivtsov, A. V. et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).
13. Bernt, K. M. et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66–78 (2011).
14. Wang, X., Chen, C. W. & Armstrong, S. A. The role of DOT1L in the maintenance of leukemia gene expression. *Curr. Opin. Genet. Dev.* **36**, 68–72 (2016).
15. Pina, C., May, G., Soneji, S., Hong, D. & Enver, T. MLLT3 regulates early human erythroid and megakaryocytic cell fate. *Cell Stem Cell* **2**, 264–273 (2008).
16. McKinney-Freeman, S. et al. The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell* **11**, 701–714 (2012).
17. Ng, E. S. et al. Differentiation of human embryonic stem cells to HOXA⁺ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nat. Biotechnol.* **34**, 1168–1179 (2016).
18. Boitano, A. E. et al. Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science* **329**, 1345–1348 (2010).
19. Fares, I. et al. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science* **345**, 1509–1512 (2014).
20. Daigle, S. R. et al. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* **122**, 1017–1025 (2013).
21. Horton, S. J. et al. MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. *Leukemia* **27**, 1116–1126 (2013).
22. Sontakke, P. et al. Modeling BCR-ABL and MLL-AF9 leukemia in a human bone marrow-like scaffold-based xenograft model. *Leukemia* **30**, 2064–2073 (2016).
23. Wagner, J. E. Jr et al. Phase I/II trial of stemregen-1 expanded umbilical cord blood hematopoietic stem cells supports testing as a stand-alone graft. *Cell Stem Cell* **18**, 144–155 (2016).
24. Ngom, M. et al. UM171 enhances lentiviral gene transfer and recovery of primitive human hematopoietic cells. *Mol. Ther. Methods Clin. Dev.* **10**, 156–164 (2018).
25. Wilkinson, A. C. et al. Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. *Nature* **571**, 117–121 (2019).
26. Kataoka, K. et al. Evf1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. *J. Exp. Med.* **208**, 2403–2416 (2011).
27. Rentas, S. et al. Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. *Nature* **532**, 508–511 (2016).
28. Komorowska, K. et al. Hepatic leukemia factor maintains quiescence of hematopoietic stem cells and protects the stem cell pool during regeneration. *Cell Reports* **21**, 3514–3523 (2017).
29. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Human haematopoietic tissue collection and processing

Human HSPCs from fetal liver and cord blood were used to assess the function of *MLLT3* in fetal and neonatal human HSC. Second trimester (14–18 weeks) fetal livers were de-identified, discarded material obtained from elective terminations of pregnancy after informed consent. Specimen age is denoted as developmental age, two weeks less than gestational age, and was determined by ultrasound or estimated by the date of the last menstrual period. Cord blood units were obtained from full-term pregnancies following informed consent and de-identified upon collection. Because these tissues are discarded material with no personal identifiers, this research does not constitute human subjects research.

Fetal liver samples were mechanically dissociated using scalpels and syringes and strained through a 70- μ m mesh. Single-cell suspensions from fetal liver and cord blood were enriched in mononuclear cells by layering on Lymphoprep (Stem Cells Technologies) and centrifugation following manufacturer instructions. $CD34^+$ cells were magnetically isolated from mononuclear cell fraction using anti- $CD34$ microbeads (Miltenyi Biotec).

OP9M2 stroma co-culture for HSPC expansion

Human HSC co-culture on OP9M2 (subclone of OP9⁺, derived in our laboratory and validated by gene expression analysis) stroma was used to test the maintenance and expansion of immunophenotypic HSPCs. OP9M2 cells were irradiated (20 Gy) and pre-plated (50,000 cells cm^{-2}) 24 h before the start of co-culture in OP9 medium, which includes α -MEM (Invitrogen), 20% fetal bovine serum (FBS; Omega) and 1 \times penicillin/streptomycin/glutamine (P/S/G). Human HSPCs were plated on stromal layer in OP9 medium supplemented with human HSC cytokines SCF (25 ng ml^{-1} , Peprotech or Invitrogen), FLT3-L (25 ng ml^{-1} , Peprotech) and TPO (25 ng ml^{-1} , Peprotech) (HSC medium). Cells were co-cultured at 37 °C and 5% CO_2 and re-plated or analysed/sorted by flow cytometry for HSPC markers ($CD34^+CD38^{-/lo}CD90^+GPI80^+$) every 7–14 days. Half of the HSC-medium was replaced every 2–3 days. Where indicated, 500 nM StemRegenerin1 (SR1) and 35 nM UMI71 (SCT) were added to the cultures to improve human HSPC expansion.

SFEM culture for HSPC expansion

To test HSPC function in clinically relevant serum-free, stroma-free culture conditions, HSPCs were plated in StemSpan SFEM II (SCT) supplemented with human SCF (100 ng ml^{-1}), human FLT3-L (100 ng ml^{-1}), human TPO (50 ng ml^{-1}), human low-density lipoprotein (10 μ g ml^{-1} , SCT), P/S/G, 500 nM SR1 and 35 nM UMI71. Cells were cultured at 37 °C and 5% CO_2 and re-plated or analysed/sorted by flow cytometry for human HSPC markers every 7–14 days. Half of the HSC medium was replaced every other day. In some experiments, DOT1L inhibitor EPZ5676 (Cayman) was added at a concentration of 500 nM.

Flow cytometry and cell sorting

FACS analysis was performed using single-cell suspensions prepared from fetal liver and cord blood as described above, or obtained from NSG mice transplanted with human cells.

For identification of human HSPCs, cells were stained with mouse anti-human monoclonal antibodies against human $CD34$ -APC cl. 581 (555824; BD, 1:20) or -BV605 (343529; Biolegend, 1:20), $CD90$ -FITC cl. 5E10 (555595; BD, 1:100) or -APC (555595; BD, 1:100), $CD38$ -PE-Cy7 or -BUV496cl. HIT2 (560677 and 564657; BD, 1:100), $GPI-80$ -PE cl. 3H9 (D087-5; MBL, used at 1:50). Cells were assayed on a BD-LSRII flow

cytometer and data were analysed with FlowJo software (Tree Star). Cell sorting was performed using a BD FACS Aria II.

To monitor human haematopoietic engraftment *in vivo* in NSG mice, bone marrow cells were stained with rat anti-mouse- $CD45$ -APC-H7 cl. 30-F11 (557659; BD, 1:100) and mouse anti-human monoclonal antibodies against human- $CD45$ -BV711 or -BV785 cl. HI30 (304050, 304048; Biolegend, 1:100). Haematopoietic differentiation was assessed *in vitro* and *in vivo* using mouse anti-human monoclonal antibodies against human $CD19$ -PE or BV605 cl. 1D3 or HIB19 (12-0193, 12-0199; eBiosciences, 1:50), $CD3$ -PE-Cy7 cl. SK7 (557851, BD) (eBiosciences, 1:50), $CD4$ -APC (MHCD0405; Invitrogen, 1:50), $CD8$ -PE cl. HIT8A (555635; BD, 1:50), $CD13$ -APC cl. WM15 (557454; BD, 1:50), $CD66b$ -BV421 cl. G10F5 (555724; BD, 1:50), $CD14$ -V500 cl. WM53 (561816; BD, 1:100), $CD235a$ -PE or -APC (Glycophorin A, HIR2, BD 1:100), $CD71$ -AF647 (DF1513, Santa Cruz 1:100), $CD41a$ -APC cl. HIP8 (579777; BD, 1:20), $CD42b$ -PEcy7 cl. HIP1 (303916; Biolegend, 1:100), $CD10$ -AF700 cl. HI10a (563509; BD, 1:100), $CD33$ -PE cl. WM53 (555450; BD, 1:50), $CD24$ -BV711 cl. ML5 (563401; BD, 1:100), $CD20$ -BV650 cl. 2H7 (563780; BD, 1:100).

Dead cells were excluded with 7AAD (BD Biosciences, used at 1:50). For assessment of cell-cycle stages or apoptosis, 7AAD was combined with BrdU-PE (556029; BD, 1:100) or annexinV-PE (556422; BD, 1:50), respectively (see below).

RNA isolation, cDNA synthesis and qRT-PCR

RNA isolation was performed using the RNeasy Mini kit (Qiagen) with additional DNase step using manufacturer's protocol. cDNAs were prepared using High-Capacity cDNA Reverse Transcription Kit (ThermoFisher), and qPCR for *GAPDH*, *MLLT3*, *MECOM* and *HLF* was performed with the LightCycler 480 SYBR Green I Master Mix (Roche) or TaqMan Gene Expression Master Mix on the Lightcycler 480 (Roche). Primers are presented in Supplementary Table 5.

Production of lentiviral shRNA and overexpression vectors

shRNA experiments were performed with pLKO lentiviral vectors from the TRC library containing puromycin resistance gene. shRNA TRCN0000005793 (93, Sigma) was selected for *in vitro* and *in vivo* experiments after testing for knockdown efficacy compared to other shRNAs against *MLLT3* from the RNAi Consortium (TRCN0000005790 to TRCN0000005794) series. TRCN0000014790 and TRCN0000002528 were used to knockdown *HLF* and *MECOM*, respectively. shRNAs were tested on the K562 and Hep2G cell lines, obtained by ATCC and mycoplasma-free.

Human *MLLT3* was cloned from human FL-HSPC full-length cDNA into the constitutive FUGW (Addgene plasmid 14883, from D. Baltimore) lentiviral vector, in which its expression is controlled by the *UBC* promoter. *MLLT3* cDNA with a C-terminal V5-tag was inserted downstream and in frame with the GFP sequence with the synthetic addition a P2A sequence between the 2 ORFs, through two rounds of PCR (Supplementary Table 4) using PfuUltra II Ultra Fusion HS (Agilent).

For lentiviral vector production, 20 million 293T cells were co-transfected with deltaR8.2 packaging plasmid, VSVG-envelope plasmid, the lentiviral vector plasmid of choice and Turbo DNAfectin 3000 (Lamda), following manufacturer's instructions, in Opti-MEM (Life Technologies) and incubated for 5–6 h at 37 °C. After incubation for 48 h in Ultraculture medium (Lonza), supernatant was filtered and concentrated by ultracentrifugation and pelleted viruses were resuspended in SFEM and stored at –80 °C.

Lentiviral transduction

Sorted fetal liver and cord blood HSPCs were prestimulated 24 h in StemSpan SFEM II supplemented with SCF, FLT3-L, TPO and antibiotic-antimycotic (LT). Wells were pre-coated with 40 μ g ml^{-1} RetroNectin (Takara) and seeded with pre-stimulated HSPC in 300 μ l SFEM. Lentivirus was added twice, first after 24 h from cell seeding and again after additional 8 h. After 48 h from cell seeding, transduced cells

Article

were washed and seeded in the indicated conditions for each assay. For shRNA-lentiviral vectors, puromycin ($1.0 \mu\text{g ml}^{-1}$) treatment was used for selection of transduced HSPCs and maintained throughout culture. Cells infected with overexpression vectors were selected by cell sorting for GFP expression.

Haematopoietic differentiation assays

FL-HSPCs sorted directly or after 4 weeks of expansion on OP9M2 were plated on differentiation assays. For myeloid assay, FL-HSPCs were plated in StemSpan SFEM II (SCT) supplemented with G-CSF (20 ng ml^{-1} , Peprotech), GM-CSF (20 ng ml^{-1} , Peprotech), TPO (25 ng ml^{-1}) and SCF (25 ng ml^{-1}). For the erythroid assay, HSPCs were plated in StemSpan SFEM II supplemented with IL-3 (20 ng ml^{-1} , Peprotech), EPO (2.5 IU , Thermo Fisher Scientific), SCF (25 ng ml^{-1}), L-glutamine and antibiotic-antimycotic. For the megakaryocyte differentiation assay, HSPCs were plated in IMDM supplemented with recombinant (0.4% Albumedix), 2-mercaptoethanol ($100 \mu\text{M}$, Gibco), SCF (100 ng ml^{-1}), TPO (50 ng ml^{-1}), IL-3 (10 ng ml^{-1}), IL-6 (7.5 ng ml^{-1} , Peprotech), IL-9 (13.5 ng ml^{-1} , Humanzyme), human low-density lipoprotein ($4 \mu\text{g ml}^{-1}$, SCT), ITS-X (Gibco), glutamax and antibiotic-antimycotic. For the T cell assay, HSPCs were plated on non-irradiated OP9-DLL1 stroma ($25,000 \text{ cells cm}^{-2}$) in OP9 medium supplemented with SCF (25 ng ml^{-1}), FLT3-L (10 ng ml^{-1}) and IL-7 (20 ng ml^{-1} , Peprotech). For the B cell assay, HSPCs were seeded on irradiated OP9M2 in MEM- α 5% FBS supplemented with SCF (25 ng ml^{-1}), FLT3-L (10 ng ml^{-1}) and IL-7 (20 ng ml^{-1}). In each assay, cells were cultured in 24-well plates with 1 ml of the indicated medium and half medium changes were applied every 2–3 days. Cells were analysed by flow cytometry after 2 weeks.

Cell cycle and apoptosis assays

For the BrdU incorporation analysis in vitro, control- or MLLT3-OE-transduced FL-HSPCs were cultured for 4 weeks on OP9M2 with 500 nM SR1 and 35 nM UMI71 and pulse-labelled with $10 \mu\text{M}$ BrdU for 35 min in culture. Cells were sorted for the indicated surface phenotypes and processed according to the PE-BrdU flow kit (BD) instructions and analysed by flow cytometry. Apoptosis from FL-HSPCs cultured in the same conditions was assessed by flow cytometry using annexin-V-PE (BD) and 7AAD incorporation, following manufacturer instructions.

Transplantation assays in NSG mice

In vivo reconstitution ability of cultured HSPCs was assessed in immunodeficient mice. Female NSG (Jackson Laboratories) mice, 8–12 weeks old, were sub-lethally irradiated (2.75 Gy) and retro-orbitally injected with cells derived from fetal liver or cord blood in a volume of $100 \mu\text{l}$ of RPMI.

For MLLT3 shRNA knockdown experiments, 30,000 fetal liver CD34⁺ cells were infected with MLLT3 shRNA or control lentivirus and cultured on OP9M2 for 9 days under puromycin selection before transplantation. At 12 weeks, MLLT3-KD mice were euthanized to obtain bone marrow.

For MLLT3-OE experiments with FL-HSPCs, mice were transplanted with the progeny of 1,000 sorted GPI80⁺ FL-HSPCs (CD34⁺CD38^{−/lo}CD90⁺GPI80⁺). Cells were transduced with the MLLT3-OE or control vector and re-sorted after 5 days for GFP⁺CD34⁺ cells. These cells were either injected retro-orbitally immediately after sorting (day 5) or expanded on SFEM for an additional 10 days (day 15) before injection. After 24 weeks, mice blood was collected, and mice were euthanized to obtain bone marrow and spleen. Collected cells were analysed by FACS to evaluate human engraftment (human CD45 and absence of mouse CD45). Reconstituted mice were defined by the presence of GFP⁺ (MLLT3-OE and control vector) human CD45 cells that differentiated to myeloid and lymphoid lineages 24 weeks after transplantation. Differentiation into myelo-lymphoid lineages was evaluated by the detection of myeloid (CD14 or CD66b), B-lymphoid (CD19) and T-lymphoid (CD3, CD4 and CD8) markers on human CD45⁺ cells. Of note, T-lymphoid engraftment was evaluated only at 24 weeks from transplantation,

owing to the absence of reliable T-lymphoid differentiation at week 12 (MLLT3-KD, Extended Data Fig. 1k). Preservation of the HSPC compartment (CD34⁺CD38^{−/lo}) was also recorded.

For BrdU incorporation analysis in vivo, NSG mice were transplanted with high doses of human FL-HSPCs (8,000 GPI80⁺ HSPCs or their progeny at culture day 15 after transduction with control or MLLT3-OE vector). Then 14 weeks after transplantation, 2 mg of BrdU was injected intraperitoneally. Mice were euthanized 100 min after and bone marrow was collected. Cells were sorted for the indicated surface phenotypes and processed according to the PE-BrdU flow kit (BD) instructions and assessed by flow cytometry.

For secondary transplantations, viably frozen bone marrow from primary NSG mice transplanted with 15-day cultured HSPCs or uncultured HSPCs was used. Bone marrow from mice from the same experimental group was pooled and a dose equivalent of 1/2 femur of the primary mouse total bone marrow was injected retro-orbitally in sub-lethally irradiated secondary recipients. Human engraftment was assessed in bone marrow after 18 weeks from transplantation.

For the limiting dilution assay, CB-HSPCs were injected after FACS sorting, after transduction and sorting transduced cells (GFP⁺CD34⁺, day 5), or after an additional 10 days of expansion in culture (day 15). The number of cells reported (2,000, 500, 50, 10) is the number of sorted HSPCs (uncultured) that was transplanted directly, or the progeny of which was transplanted after culture (days 5 and 15). The criteria for multilineage engraftment was having hCD45 positive (GFP⁺ with MLLT3 and CTR vector) cells that displayed at least one myeloid (CD14 CD66b or both) and the B-lymphoid (CD19) marker. HSC frequency was assessed using ELDA software²⁹.

A panel of markers covering the classical AML blast (CD34 and CD33) and B-cell precursor markers (CD34, CD10, CD20, CD24, CD38) altered in MLL-MLLT3 fusion gene-driven leukaemia in immunodeficient mice were tested in mice transplanted with day-15-expanded MLLT3-OE CB-HSPCs and uncultured CB-HSPCs.

All transplanted mice were included in the analysis unless they died before the experimental endpoint (in total from combined fetal liver and cord blood experiments, 14 out of 115 mice transplanted with MLLT3-OE, 17 out of 98 mice transplanted with control vector HSPCs, and 3 out of 36 mice with uncultured HSPCs died before 24 weeks).

All studies and procedures involving mice were conducted in compliance with all the relevant ethical regulations and were approved by the UCLA Animal Research Committee (protocol 2005-109).

ChIP-seq analysis

Sorted FL-HSPCs or erythroblasts (CD34⁺CD235⁺CD71⁺) (50,000–10,000 per immunoprecipitation) were crosslinked in 1% formaldehyde for 10 min, quenched with glycine 0.0125 M and snap-frozen as a dry pellet. The pellet was re-suspended in lysis buffer (50 mM Tris, pH 8.2, 10 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.5% sarkosyl) and sonicated 12 min at a 5% intensity using Misonix cup-horn sonicator. Chromatin was incubated overnight with $2 \mu\text{g}$ of antibody (anti-MLLT3, Genetex, GTX102835; anti-V5, Abcam, ab15828; H3K4me3, Abcam ab8580; H3K9ac, CST 9649; H3K9cr, PTMPTM-516; H3K36me3, Abcam ab9050; H3K79me2, Abcam, ab3594; RNA-polIII Rbp1 (8WG16) Biolegend 664911, H3K27ac, AM 39133; H3K4me1, AM 39297; H3K27me3, AM 39155; H3K9m3, AM 39161) preloaded onto $20 \mu\text{l}$ of Protein G Dynabeads (Thermo Fisher) and washed twice with each of the solutions (low-salt wash, high-salt wash, LiCl wash and TE buffer) as previously described³⁰.

ChIP experiments for H3K79me2 were also performed from short-term cultures with FL-HSPCs transduced with either MLLT3-KD (no. 93) or MLLT3-OE vector and respective controls. Knockdown was collected at day 5, which included 3 days of puromycin selection. MLLT3-OE were sorted for GFP and collected at day 10, which included 4 days of EPZ5676 treatment at 500 nM, or DMSO as a control. Sonicated chromatin from these chips was spiked-in with S2 chromatin (Activ Motif)

following manufacturer instructions. Libraries were prepared with the Nugen Ovation Ultralow kit v2 following manufacturer instructions and sequenced using HiSeq-4000 (Illumina) to obtain single-end 50-bp long reads. Demultiplexing of the reads based on the barcoding was performed using in house Unix shell script. Mapping to the human genome (hg19) was performed using bowtie2. Samtools v.1.3.1 package was used to create a .bam file, remove duplicates, blacklisted region and ChrM regions, sort and index. Bedtools multicov was used to quantify histone mark signals. GO analysis on genes bound by MLLT3 was calculated using Homer annotatePeaks.pl³¹ using -go function and statistic is reported as *P* value, calculated with standard parameters by the algorithm. Each H3K79me2, H3K4me3, H3K36me3 and H3K9ac ChIP replicate on MLLT3-OE (Fig. 3f) is internally normalized using the non-MLLT3-bound, housekeeping gene *VCL* (Extended Data Fig. 8b). For H3K79me2 ChIP samples spiked in with S2 chromatin, sequences were aligned to both the human and *Drosophila melanogaster* genome (dm6) and the quantified reads, were normalized by total reads aligned to dm6 for each sample (Fig. 3g). Coverage files, average profiles heat maps were created with Deeptools packages³². MACS2 v.2.1.1³³ was used to call MLLT3 peaks using default parameters for broadPeak calling.

ATAC-seq analysis

FL-HSPCs or erythroblasts (CD34⁺CD235⁺CD71⁺) (50,000 sorted cells) were processed according to the protocol³⁴, with minor adjustments. Nuclei were purified by the addition of 250 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630) to sorted cells, pelleted and resuspended in the transposition reaction mix (Nextera DNA Library Prep Kit, Illumina) and incubated at 37 °C for 30 min. Transposed DNA was column purified and used for library amplification with custom made adaptor primers³⁴ using NEBNext High-Fidelity 2× PCR Master Mix (New England Labs). The amplification was interrupted after 5 cycles and a SyBR green qPCR was performed with 1/10 of the sample to estimate for each sample the additional number of cycles to perform before saturation was achieved. Total amplification was between 10 and 15 cycles. Purified libraries were sequenced using HiSeq-2000 (Illumina) to obtain paired-end 50-bp long reads. Read-mapping to the genome (hg19) was done using Bowtie2 or v.2.2.9³⁵ with parameters—local -X 2000 -N 1—no-mixed. The Bamcoverage tool from Deeptools was used to create the coverage .bw files for visualization³². Samtools v.1.3.1 was used to remove duplicates and reads aligned to chrM.

RNA-seq analysis

Total RNA from 50,000 sorted HSPCs was extracted using the RNeasy Mini kit (Qiagen) and library was constructed using KAPA RNA Hyper-Prep Kit with RiboErase (HMR). Libraries were sequenced using HiSeq-4000 (Illumina) to obtain paired-end 50-bp long reads. Mapping to the human genome (hg19) was performed using TopHat v.2.0.9 or v.2.0.14³⁶ with the parameters—no-coverage-search -M -T -x 1. Coverage files were created with the Genomecov tool from Bedtools³⁷ with the parameters -bg -split -ibam. For abundance estimations (FPKMs) the aligned read files were further processed with HOMER coupled to edgeR on the hg19 annotation. GO was calculated using DAVID³⁸ and statistic is reported as *P* value, calculated with standard parameters by the algorithm. Gene expression changes were considered significantly up- and downregulated when adjusted *P* < 0.05 (−log-transformed fold change > 0.322). The selection of differentially expressed genes is based on adjusted *P* value (significant across six replicates) rather than fold change because, as MLLT3 is required to maintain HSC identity and

viability, the changes observed in the HSPC compartment reflect the beginning of MLLT3-dependent processes, not the end point. Heat maps were generated using Morpheus (Broad Institute) using a colour scale normalized by the minimum and maximum value for each gene.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence data that support the findings of this study have been deposited in Gene Expression Omnibus (GEO) with the accession code GSE111484. Data from published reference are available in GEO GSE81080¹⁷. All other data are either available within the paper or from the corresponding author upon reasonable request. Custom codes for data analysis are also available upon request. There is no restriction in data availability.

30. Org, T. et al. Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. *EMBO J.* **34**, 759–777 (2015).
31. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
32. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91 (2014).
33. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
34. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
36. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
37. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
38. Dennis, G. Jr et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).

Acknowledgements We thank BSCRC flow cytometry (F. Codrea, J. Scholes and J. Calimlim) and sequencing (S. Feng) cores, TCGB (NIH P30CA016042) and CFAR cores (NIH P30AI028697-21) at UCLA. We thank J. Zhao, T. Montoya and D. Dou for assistance with experiments; V. Rezek, D. Johnson and O. Witte for help with NSG mice and D. Kohn, Z. Romero and R. Hollis for help with lentiviral vectors. This work was supported by NIH RO1 DK100959 and RO1 DK121557; Broad Stem Cell Center at UCLA, Rose Hills Foundation, Jonsson Cancer Center Foundation, and UCLA David Geffen School of Medicine Regenerative Medicine Theme Award for H.K.A.M.; LLS Special Fellow Award and BSCRC post-doctoral fellow award for V.C.; CIRM GC1R-06673B for C.M.G. and H.K.A.M.; Ruth L. Kirschstein National Research Service Award T32HL069766 for A.V., LLS Fellow Award and AACR post-doctoral award for F.L.; Beckman Scholars Program for A.T.N. and T.J.B.; and NIH grant CA178415 to S.K.K.

Author contributions V.C. and H.K.A.M. designed experiments and interpreted data. V.C. performed and/or supervised all experiments and data analysis, including the generation of functional data documenting the effects of MLLT3 on HSC expansion in culture, the characterization of the expanded HSCs in transplantation and the molecular characterization of MLLT3-mediated gene regulation in HSCs. A.T.N., T.J.B. and A.V. contributed to MLLT3 knockdown and overexpression experiments. Other members of the laboratory also independently replicated these key functional data. A.V., T.J.B., L.K.L. and Y.W. contributed to the transplantation experiments in NSG mice. V.C., F.D.L. and T.S. performed bioinformatics analysis. M.M., G.M.C. and S.K.K. provided support with experimental design and discussed data analysis and results. V.C. and H.K.A.M. wrote the manuscript, which all authors edited and approved.

Competing interests The authors declare no competing interests.

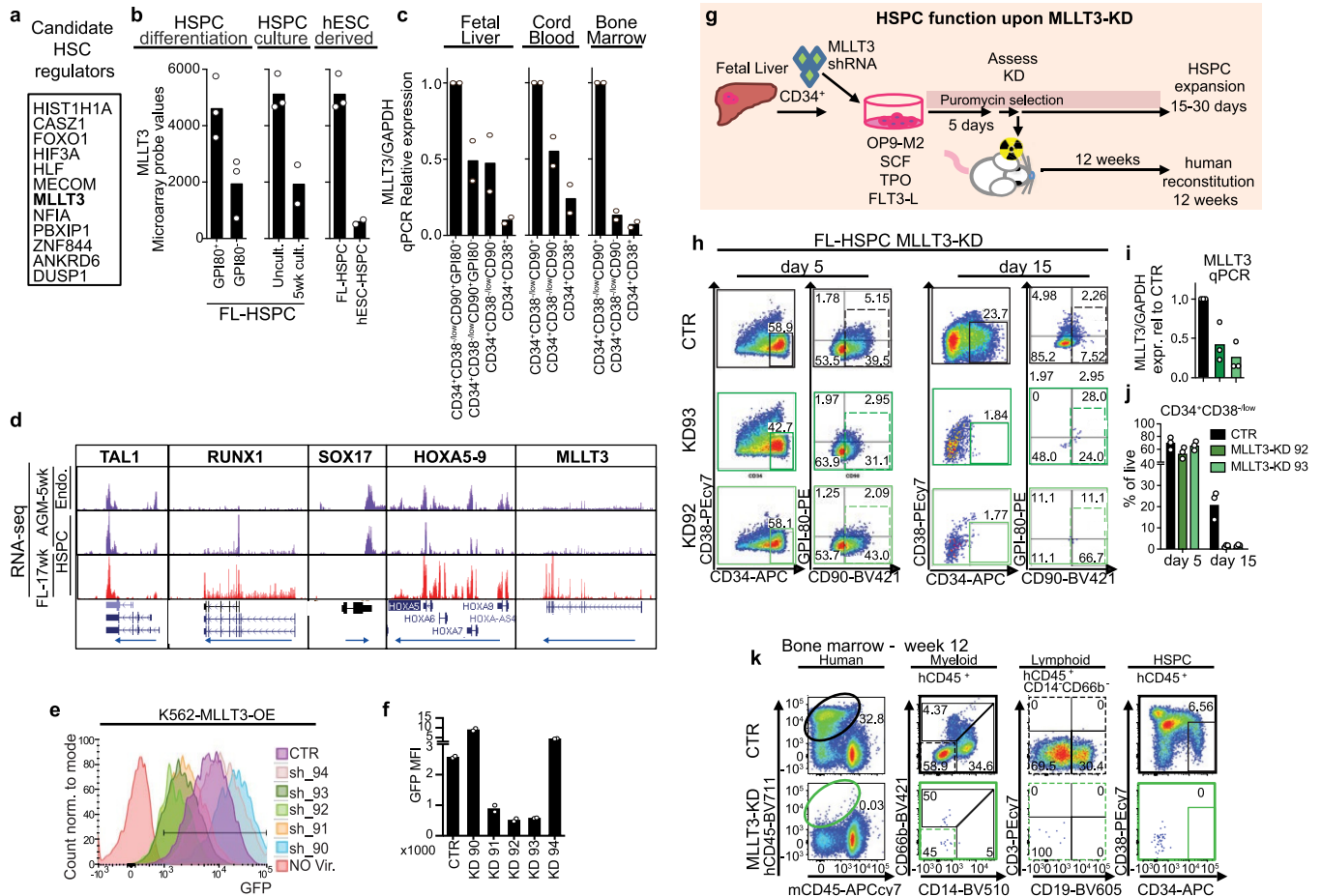
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1790-2>.

Correspondence and requests for materials should be addressed to V.C. or H.K.A.M.

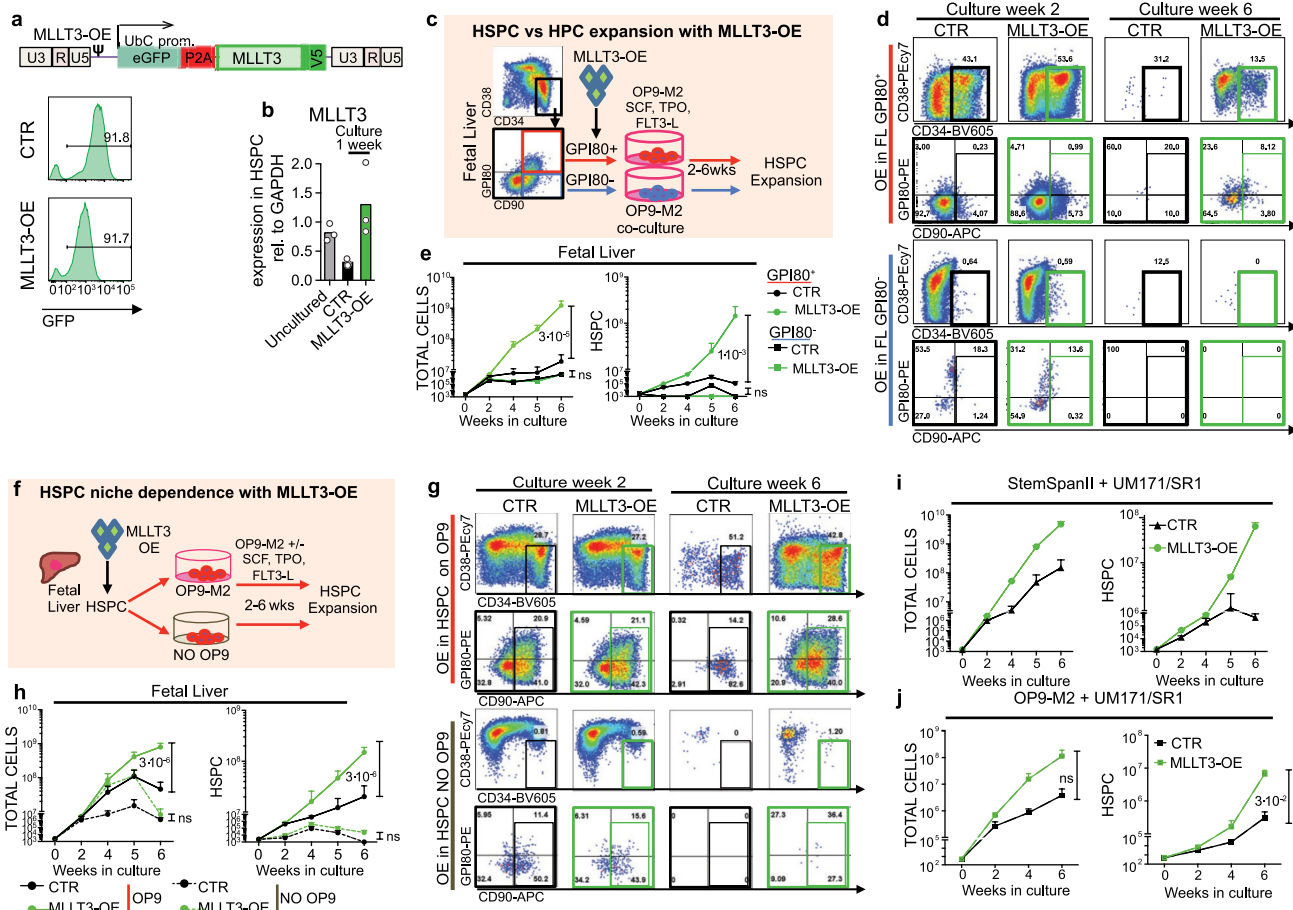
Peer review information Nature thanks Elisa Laurenti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



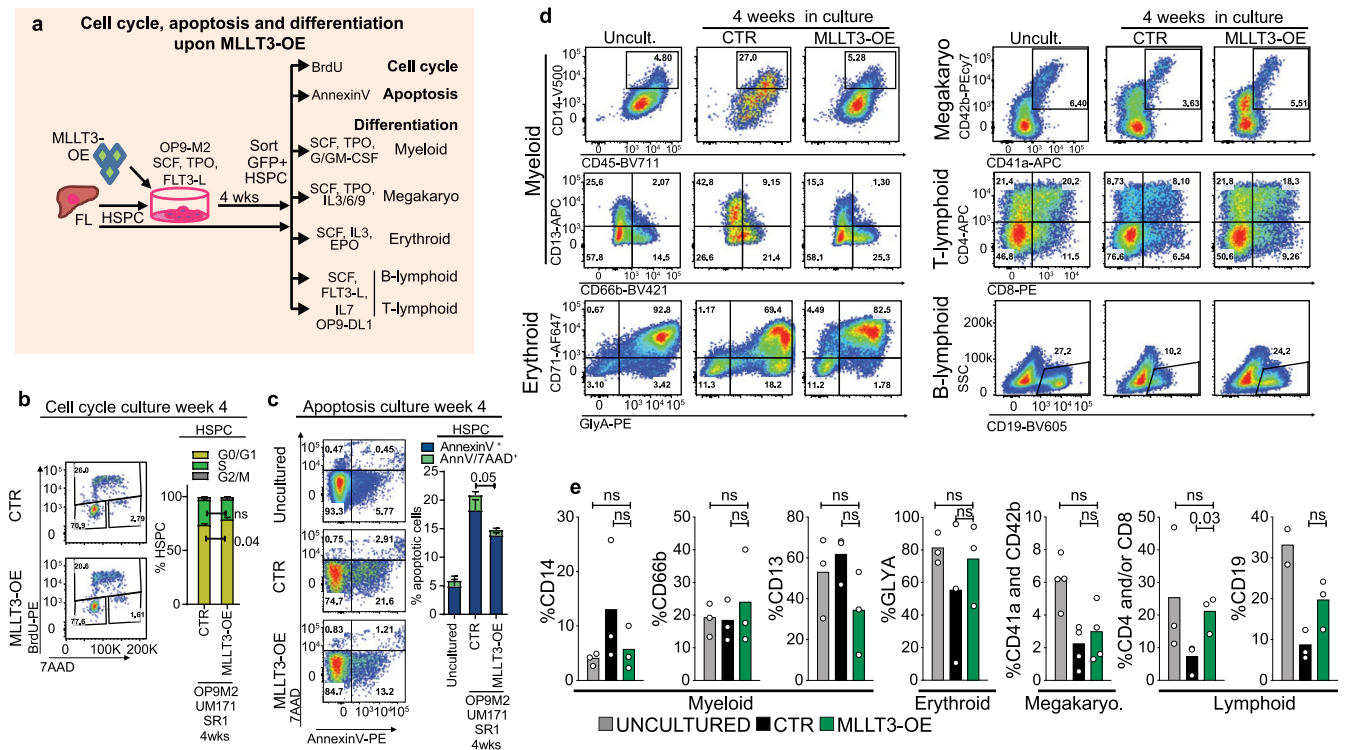
Extended Data Fig. 1 | MLLT3 expression in human haematopoietic tissues and after shRNA knockdown. **a**, Candidate human HSC self-renewal factors that correlate with the self-renewal of human FL-HSCs (Fig. 1a) and are localized in the nucleus are listed. **b**, Bar graphs showing mean and individual probe values for *MLLT3* from microarray datasets after differentiation of CD34⁺CD38^{low}CD90⁺GPI80⁺ FL-HSCs to CD34⁺CD38^{low}CD90⁺GPI80⁺ progenitors³, in FL-HSPCs after 5-week culture on OP9M2 stroma⁴, and in hES cell-HSPCs⁵ ($n = 3$; $n = 2$ for 5-week culture and hES cell-HSPCs). **c**, Quantitative PCR with reverse transcription (qRT-PCR) validation of *MLLT3* expression in human fetal liver, cord blood and bone marrow HSPCs and downstream progenitors ($n = 2$ donors per tissue). Expression is relative to *GAPDH*, normalized to the first sample of each plot. Mean and individual values are shown. **d**, UCSC genome browser tracks of RNA-seq analysis of sorted human developmental haematopoietic populations¹⁷. First trimester (5 weeks) embryonic aorta-gonad-mesonephros (AGM) endothelium (CD34⁺CD90⁺CD43⁺) and HSPC (CD34⁺CD90⁺CD43⁺) population from one donor are compared to second trimester (17-week) FL-HSPCs (CD34⁺CD38^{low}CD90⁺CD45⁺). Expression of transcription factors involved in the development

of haemogenic endothelium and HSC specification are compared to *MLLT3*. **e**, **f**, Flow cytometry analysis (**e**) and quantification (**f**) of GFP fluorescence in human K562 cells (haematopoietic cell line with no detectable *MLLT3* expression) stably expressing *MLLT3*-OE vector (as in Extended Data Fig. 2a), that were transduced with five different shRNAs targeting *MLLT3* (sh_90–94), representative of two transductions. shRNA-mediated knockdown targeting *MLLT3* also lowers GFP that is transcribed from a single GFP-P2A-*MLLT3* transcript. Median fluorescence intensity (MFI) from $n = 2$ transductions (mean and individual values). **g**, Strategy for *MLLT3* lentiviral knockdown and functional analysis in FL-HSPCs. **h**, FACS plots 5 and 15 days after *MLLT3*-KD (representative of three experiments). **i**, **j**, *MLLT3* qRT-PCR 5 days after transduction (**i**) and quantification of CD34⁺CD38^{low} (**j**) in cells transduced with empty vector or *MLLT3* shRNA (KD92 and KD93) after 5 and 15 days in culture (mean and individual values, $n = 3$ donors). **k**, Representative FACS plots from NSG mouse bone marrow 12 weeks after transplantation assessing human CD45⁺ cells, multi-lineage haematopoietic reconstitution and HSPCs (extended from Fig. 1d, $n = 10$ mice, 2 independent experiments).



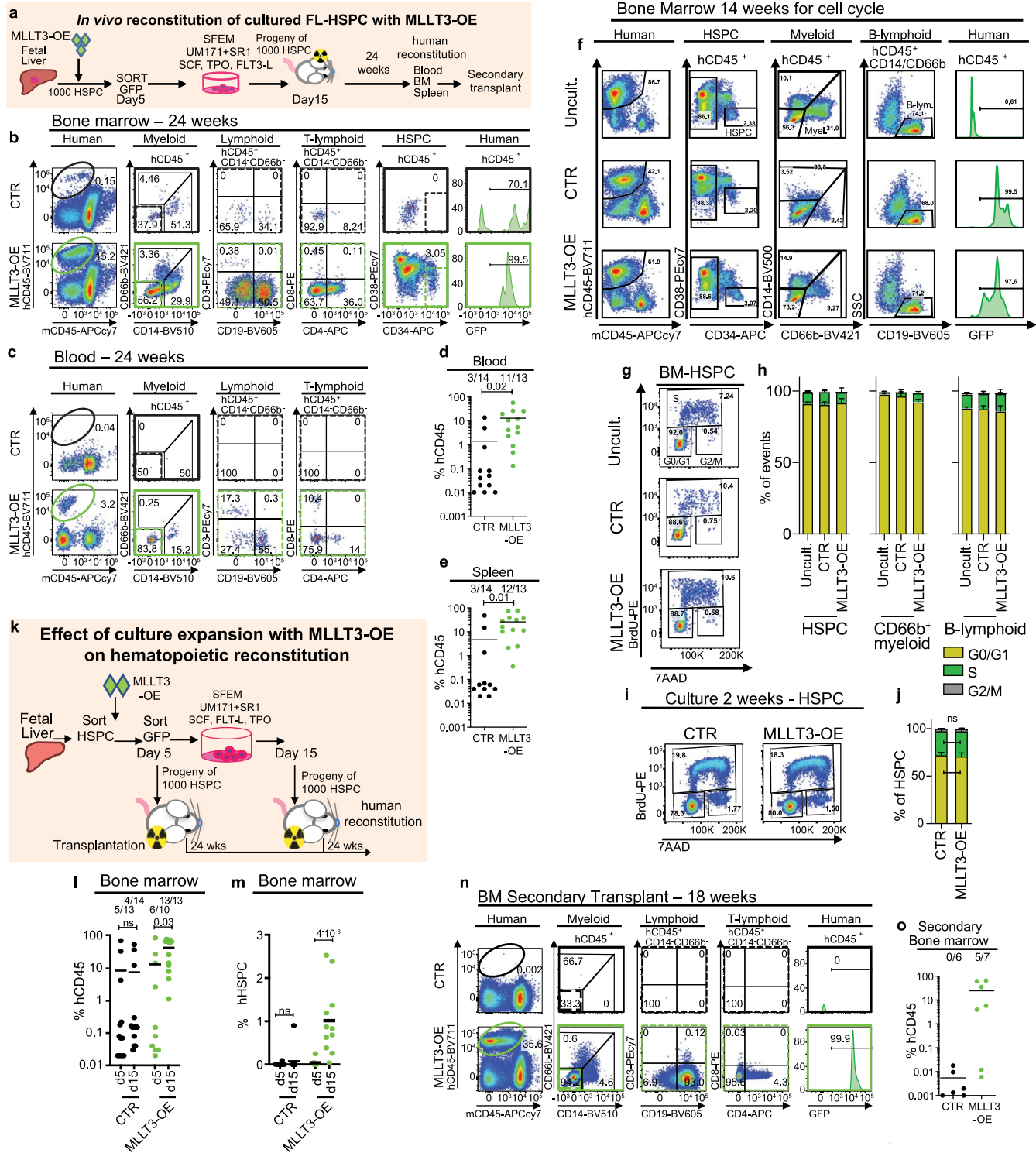
Extended Data Fig. 2 | Effects of MLLT3-OE on HSPC expansion. **a**, FUGW lentiviral construct expressing MLLT3 and GFP. P2A, cleavage sequence; UbC, ubiquitin C promoter; V5, V5 peptide tag. Representative flow cytometry plots showing persistent GFP expression in FL-HSPCs after transduction with control or MLLT3-OE vector, representative of six experiments. **b**, qRT-PCR quantifying MLLT3-OE in transduced HSPCs after one week (mean, $n = 3$). **c**, The effects of MLLT3-OE in GPI80⁺ (red) and GPI80⁻ (blue) FL-HSPC progenitors. **d**, Representative FACS plots derived from GPI80⁺ HSPCs (top) and GPI80⁻ progenitors (bottom) expressing control and MLLT3-OE vectors. **e**, Expansion of total live cells (left) and HSPCs (right) expressing control or MLLT3-OE vectors (mean and s.e.m., $n = 3$). **f**, Strategy for lentiviral overexpression of MLLT3 in FL-HSPCs seeded with or without OP9M2 stroma to assess their dependence on a supportive microenvironment. **g**, Representative FACS plots

of FUGW empty-vector-transduced and MLLT3-OE-transduced GPI80⁺ HSPCs seeded on OP9M2 (top) or stroma-free culture wells. **h**, Expansion of total live cells (left) and CD34⁺CD38⁻CD90⁺ HSPCs (right) transduced with control or MLLT3-OE vectors (mean and s.e.m., $n = 3$ independent experiments). Culture of MLLT3-OE HSPCs with small molecules UM171 and SR1. **i**, Expansion of total live cells (top) and CD34⁺CD38⁻CD90⁺ HSPCs (bottom) transduced with control or MLLT3-OE vectors (mean and s.e.m., $n = 2$ independent experiments), cultured in SFEM with SR1 and UM171. **j**, Expansion of total live cells (top) and CD34⁺CD38⁻CD90⁺ HSPCs (bottom) transduced with control or MLLT3-OE vectors (mean and s.e.m., $n = 3$ independent experiments), cultured in HSC medium on OP9M2 with SR1 and UM171. P values in **e**, **h** and **j** determined by two-sided t -test.



Extended Data Fig. 3 | Effects of MLLT3-OE on cell cycle, viability and differentiation in culture. **a**, Strategy to test cell-cycle activity, apoptosis and differentiation potential of MLLT3-OE-expanded HSPCs. **b**, BrdU incorporation flow cytometry analysis of FL-HSPCs cultured for 4 weeks on OP9M2 with SR1 and UM171. G0/G1 (BrdU⁺7AAD^{int}), S (BrdU⁺7AAD^{int}) and G2/M (BrdU⁺7AAD^{hi}) phases are quantified in FL-HSPCs (mean and s.e.m., $n = 5$). **c**, Apoptosis analysis of HSPCs cultured for 4 weeks on OP9M2 with SR1 and UM171. Early (AnnV⁺7AAD⁻) and late (AnnV⁺7AAD⁺) apoptosis are quantified by flow

cytometry (mean and s.e.m., $n = 2$ uncultured, $n = 4$ control and MLLT3-OE). **d**, FACS analysis of in vitro differentiation (representative of three experiments). **e**, Quantification of the differentiation of 4-week expanded HSPCs into myeloid, erythroid, megakaryocytic, B- and T-lymphoid lineages, compared with the differentiation of uncultured FL-HSPCs (mean, $n = 3$ donors except for megakaryocytic $n = 4$ and B-lymphoid uncultured $n = 2$). All P values determined by two-sided t -test.

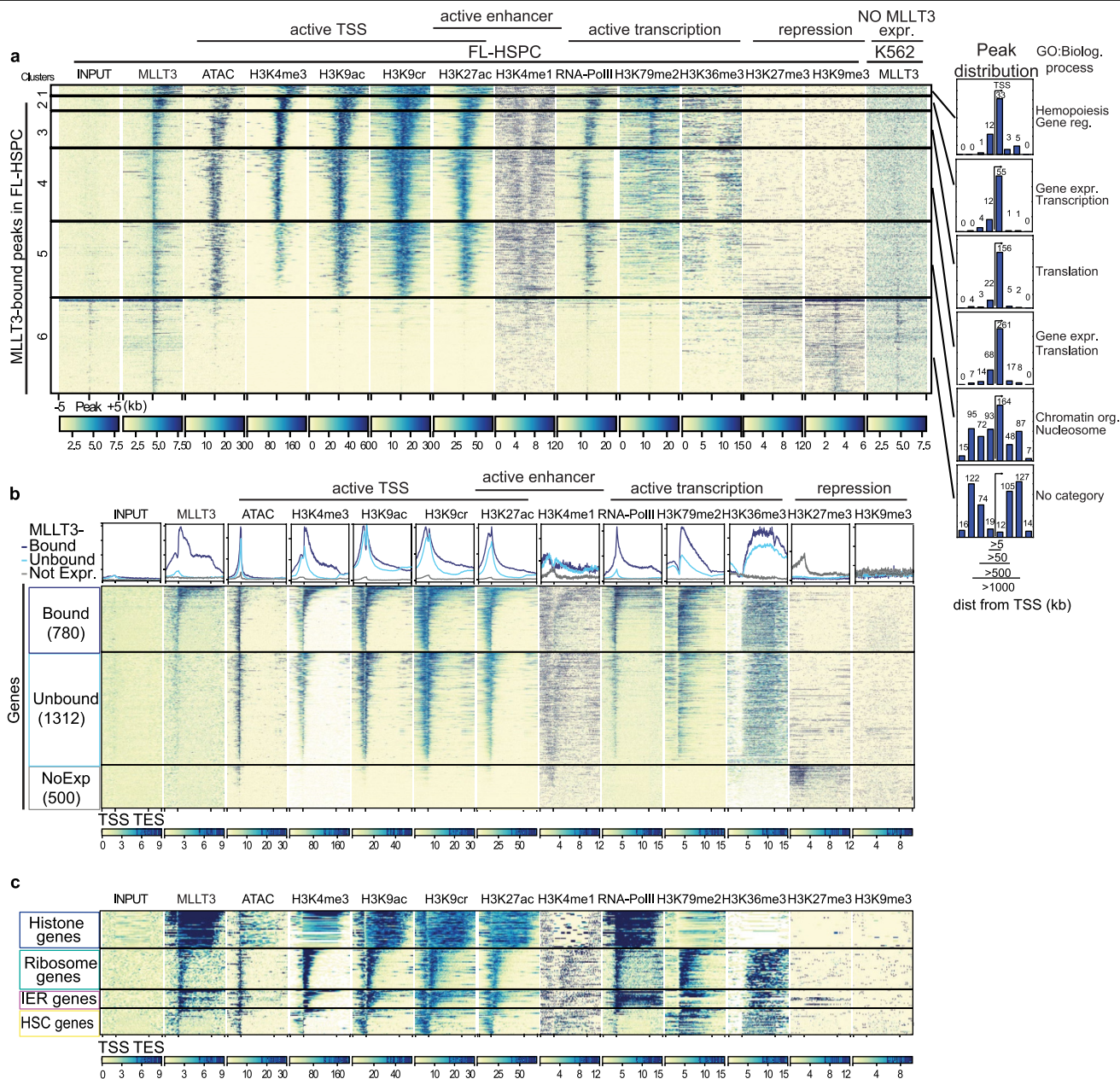


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Analysis of human reconstitution and correlation of cell cycle and culture time with the engraftment of MLLT3-OE cells.

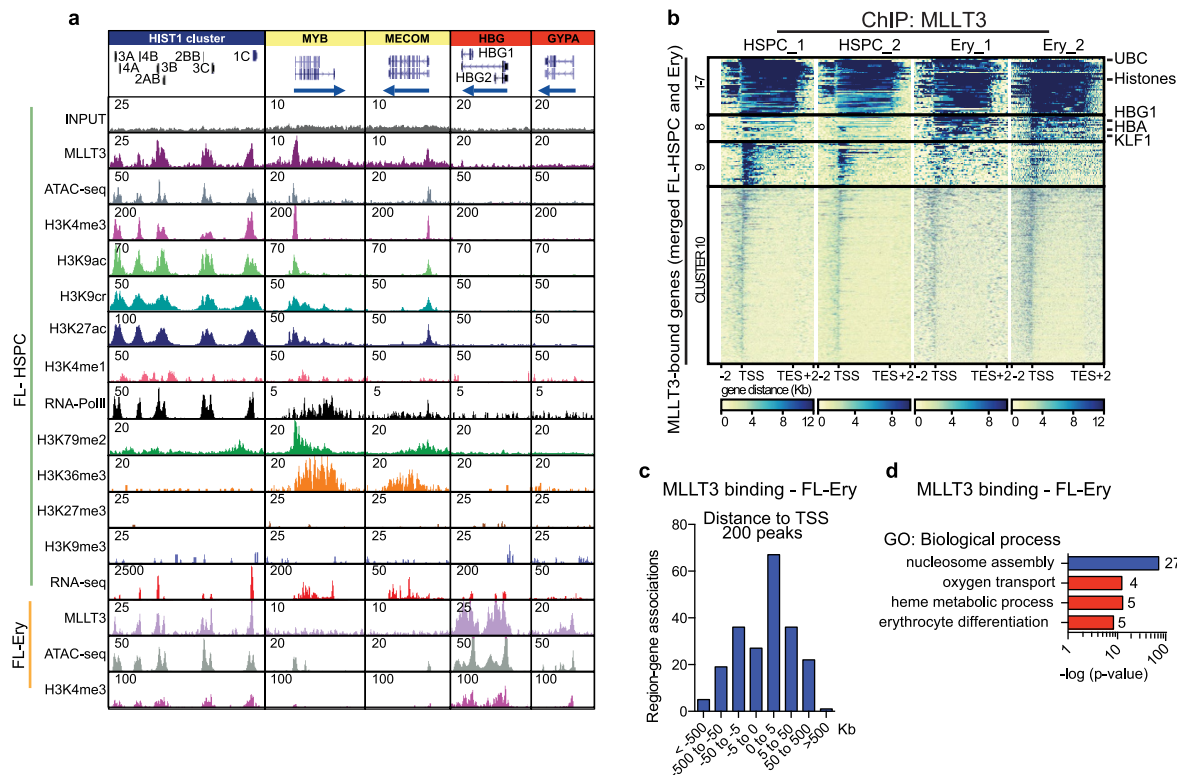
a, Transplantation strategy with ex vivo expanded FL-HSPCs transduced with control or MLLT3-OE vector. On day 5, transduced CD34⁺GFP⁺ cells derived from 1,000 HSPCs were sorted and replated in SFEM and transplanted into NSG mice at day 15. Human haematopoietic reconstitution was assessed at 24 weeks. Bone marrow from engrafted mice was transplanted to secondary recipients. **b**, FACS analysis showing human haematopoietic reconstitution (hCD45) (Fig. 1h) and multilineage differentiation (myeloid CD66, CD14; B-lymphoid CD19; T-lymphoid CD3, CD4, CD8). **c**, **d**, Representative FACS plots (**c**) and quantification (**d**) of human haematopoietic reconstitution from peripheral blood of transplanted NSG mice (extended from Fig. 1h). **e**, Quantification of human haematopoietic reconstitution from mouse spleen. Number of mice showing multi-lineage reconstitution versus total transplanted mice is shown ($n = 13$ or 14 mice, mean and individual values from four independent experiments). **f**, FACS sorting strategy for cell-cycle analysis after 100 min in vivo BrdU pulse in haematopoietic cells derived from human FL-HSPCs (8,000 GPI80⁺ FL-HSPCs or their progeny day 15 of culture after transduction with control or MLLT3-OE vector) transplanted in NSG mice and sorted 14 weeks after transplantation. **g**, **h**, Representative FACS plots (**g**) and

quantification (**h**) of cell-cycle distribution by BrdU and DNA content (7AAD) staining. G0/G1 (BrdU⁺7AAD^{lo}), S (BrdU⁺7AAD^{int}) and G2/M (BrdU⁺7AAD^{hi}) phases are quantified in HSPCs, CD66b⁺ myeloid and CD19⁺ B-lymphoid cells (mean, $n = 2$). **i**, **j**, Cell-cycle analysis in HSPCs before transplantation, after 2-week culture in SFEM with SR1 and UM171. All P values determined by two-sided t -test. **k**, Transplantation strategy assessing the effect of extended culture on the reconstitution ability of MLLT3-OE FL-HSPCs. FL-HSPCs transduced with control or MLLT3-OE vector were sorted on day 5, and CD34⁺GFP⁺ cells derived from 1,000 uncultured HSPCs were transplanted directly into NSG mice or cultured in SFEM for an additional 10 days before transplantation. Numbers show mice with multilineage human engraftment, transplanted before (day 5) and after (day 15) culture expansion. **l**, **m**, Human CD45⁺ cells (**l**) and human HSPCs (**m**) in mouse bone marrow are quantified ($n = 4$ independent experiments; number of mice is indicated). $*P < 0.05$, two-sided t -test. **n**, **o**, Representative FACS plots (**n**) and quantification (**o**) of human haematopoietic (CD45) reconstitution and multilineage differentiation in mouse bone marrow after 18 weeks of secondary transplantation (mean, two independent experiments). Secondary mice were transplanted with 2-week-old bone marrow from primary recipients (1/2 of femur of the primary transplanted mice injected per recipient).



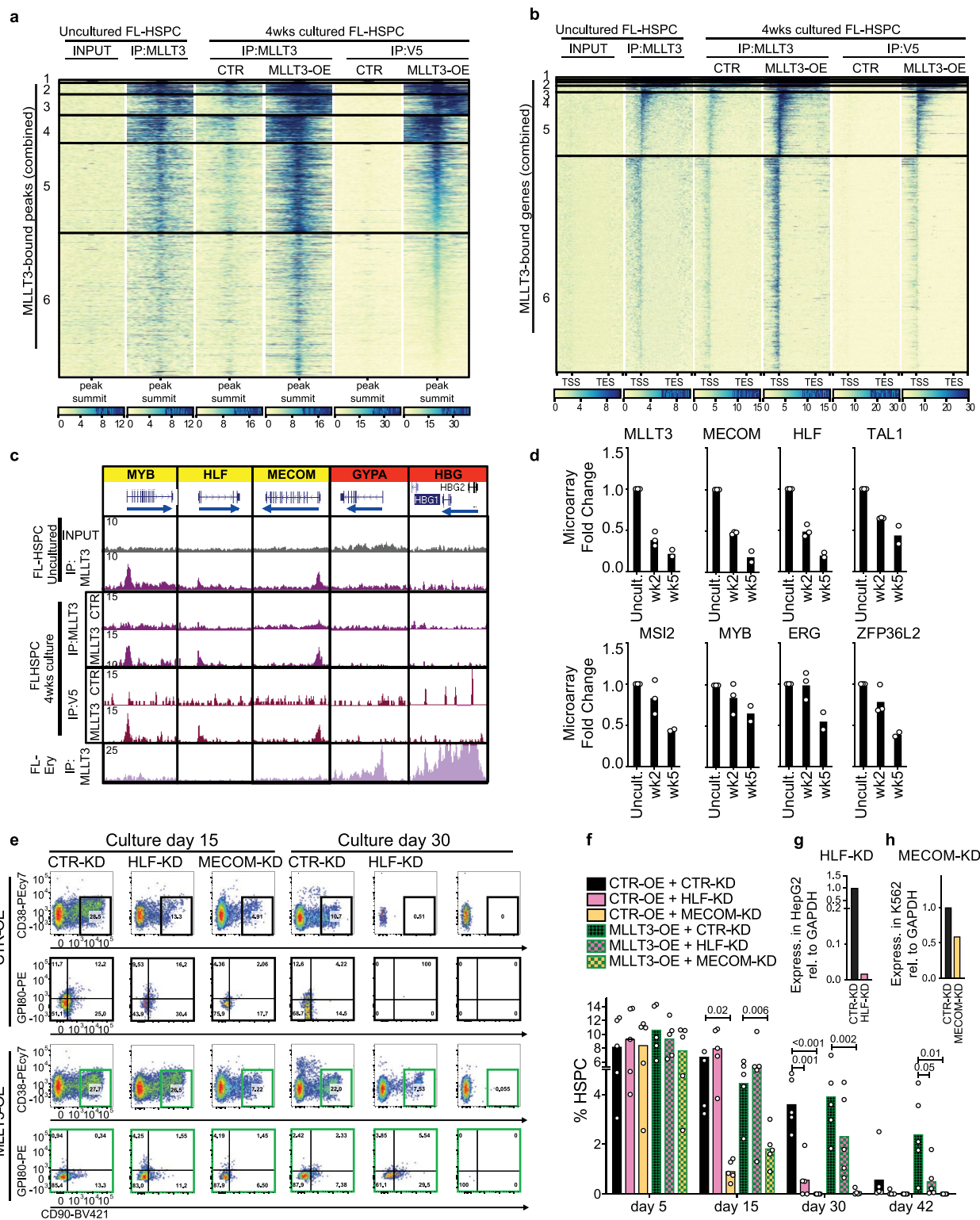
Extended Data Fig. 5 | MLLT3 binding and correlation with epigenetic marks. a, Heat map showing *k*-means clustering for MLLT3 ChIP-seq peaks and their association with ATAC-seq and histone marks in uncultured FL-HSPCs. ChIP of MLLT3 in K562 cells is shown as a control. The region surrounding the MLLT3 peak centre ± 5 kb is shown. Peak distance from the gene TSSs and representative high-ranked biological processes enriched in GREAT analysis are shown for each cluster. Cluster 6 genes, in which the MLLT3 signal does not correlate with active marks, show no significant enrichment for biological

processes. **b,** Average profile and heat map for MLLT3 and indicated histone marks in expressed and non-expressed genes in uncultured FL-HSPCs. Metagene plot ± 2 kb is shown. This represents an extended version of Fig. 2c. **c,** Heat map for selected gene groups as in **b**. This represents an extended version of Fig. 2e. Tracks in **a–c** are representative of $n = 3$ experiments for MLLT3 ChIP in FL-HSPCs, $n = 1$ for H3K9cr, H3K27me3 and H3K9me3, and $n = 2$ for other ChIP experiments.



Extended Data Fig. 6 | Cell-type specificity of MLLT3 binding. **a**, UCSC genome browser tracks showing ChIP-seq of MLLT3 and epigenetic marks in representative MLLT3-bound genes in FL-HSPCs and FL-erythroblasts ($n = 3$ MLLT3 ChIP in FL-HSPCs, $n = 1$ for H3K9cr H3K27me3 and H3K9me3, and $n = 2$ for other ChIP experiments). This represents an extended version of Fig. 2g. **b**,

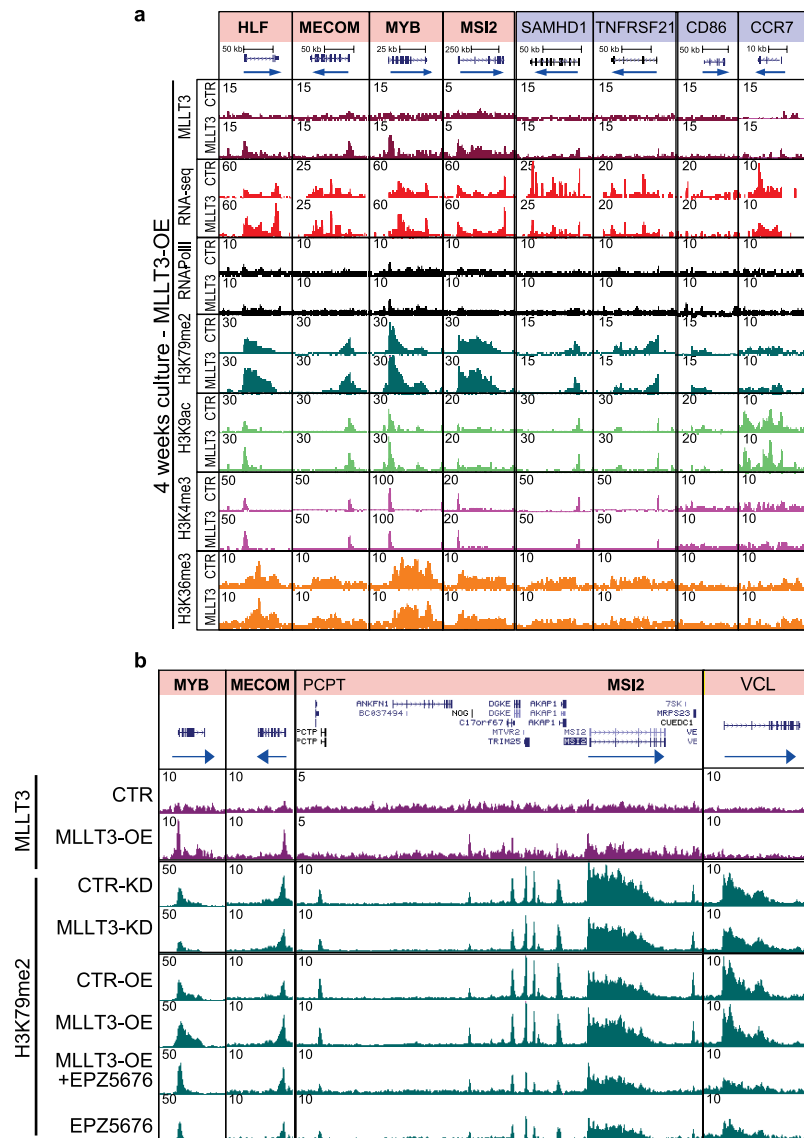
Heat map with hierarchical clustering for genes bound by MLLT3 in FL-HSPCs ($n = 2$) and/or CD34⁺GlyA⁺CD71⁺ FL-erythroblasts ($n = 2$) (genes with significant MLLT3 peak < 5 kb from TSS) showing clusters of common and cell-type-specific MLLT3-bound genes. **c**, Distribution of MLLT3 FL-erythroblast peaks. **d**, GO analysis of MLLT3-bound genes in FL-erythroblasts.



Extended Data Fig. 7 | See next page for caption.

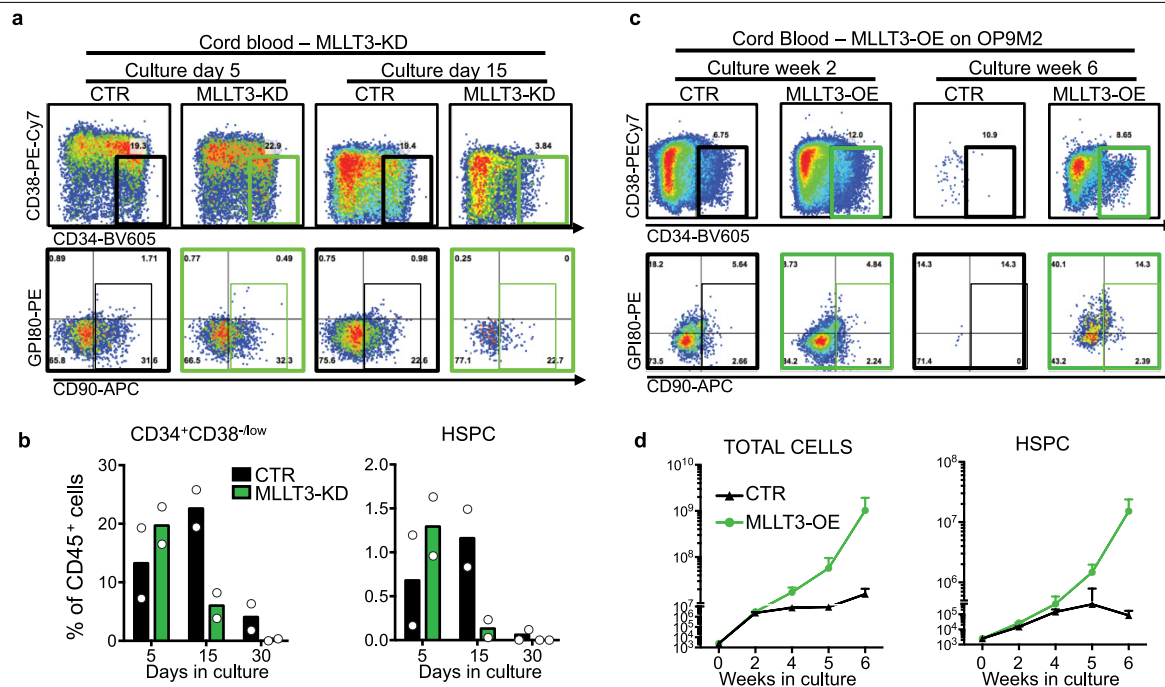
Extended Data Fig. 7 | MLLT3 binding and regulation of MLLT3 target genes in cultured HSPCs. **a**, Heat map showing distribution of MLLT3 peaks reflecting endogenous MLLT3 and MLLT3-OE ChIP-seq signal using MLLT3 and V5 antibodies. MLLT3 from uncultured FL-HSPCs and input control, MLLT3 ChIP signal from 4 week cultured FL-HSPCs with MLLT3-OE or empty control vector, and V5 ChIP from FL-HSPCs cultured for 4 weeks with MLLT3-OE or control vector. The region surrounding the peak summit (± 2 kb) is shown for the combined MLLT3 peaks from all conditions (FL-HSPC MLLT3 ChIP $n = 3$, MLLT3-OE V5 ChIP $n = 2$ and MLLT3-OE MLLT3 ChIP $n = 1$). IP, immunoprecipitate. **b**, Heat map showing hierarchical clustering for all MLLT3-bound genes as in **a**. Signal from all MLLT3-bound genes (significant MLLT3 peak at < 5 kb from TSS) is shown. **c**, UCSC genome browser tracks of representative MLLT3-bound HSC (yellow) and erythroid (red) genes. Tracks show ChIP-seq signal for input and MLLT3 in uncultured FL-HSPCs, MLLT3 ChIP-seq on FL-HSPCs cultured for 4 weeks with control and MLLT3-OE

vectors; V5-tagged MLLT3 ChIP-seq from FL-HSPCs cultured for 4 weeks with control and MLLT3-OE vectors, and MLLT3 ChIP-seq in FL-erythroblasts. **d**, Probe values from microarray datasets⁴ for MLLT3 and MLLT3-bound and upregulated HSC genes (see Fig. 3) documenting the decline of gene expression over time in culture (day 0 ($n = 3$) 2 weeks ($n = 3$) and 5 weeks ($n = 2$) on OP9M2, mean and individual values). **e–g**, Knockdown of MLLT3 targets MECOM and HLF in MLLT3-OE cells. **e**, FACS plots at 15 and 30 days after MECOM-KD and HLF-KD, with or without MLLT3-OE (representative of three experiments). **f**, Quantification of CD34⁺CD38^{−/lo}CD90⁺ HSPCs (percentage of live cells) in MECOM-KD and HLF-KD FL-HSPCs transduced with empty vector or MLLT3-OE after 5, 15, 30 and 42 days in culture (mean, $n = 3$). *P* values determined by two-sided *t*-test. **g**, Quantification of HLF downregulation in HepG2 cells by qRT-PCR. **h**, Quantification of MECOM downregulation in K562 cells by qRT-PCR.



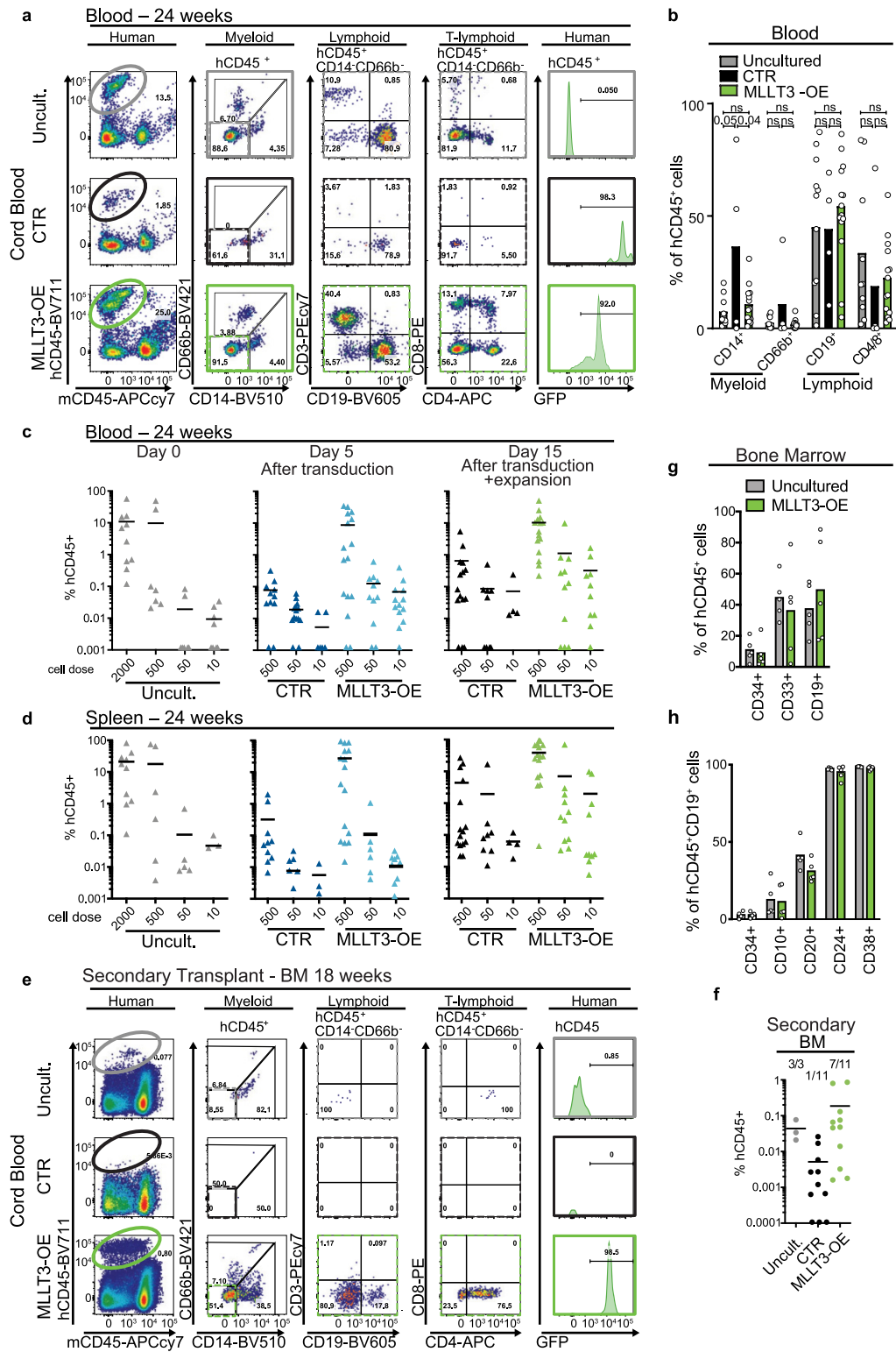
Extended Data Fig. 8 | Examples of MLLT3-mediated gene regulation after MLLT3-OE, MLLT3-KD and DOT1L inhibition. a, UCSC genome browser tracks of representative MLLT3-OE upregulated or downregulated genes. From top to bottom: MLLT3 ChIP-seq signal in FL-HSPCs cultured for 4 weeks with empty vector or MLLT3-OE; RNA-seq tracks; ChIP-seq for RNA Pol II and epigenetic marks in cultured control and MLLT3-OE FL-HSPC. Tracks are generated from the merged signal from $n = 5$ H3K79me2 ChIP, $n = 2$ other ChIPs. **b,** UCSC

genome browser tracks of the indicated ChIP experiments in representative genes in MLLT3-KD and control vector after 5-day culture, MLLT3-OE, MLLT3-OE plus EPZ5676 (DOT1L inhibitor), or control plus EPZ5676 and control vector after 10-day culture, and MLLT3-OE versus control from 4-week culture. Genes from Fig. 3g are indicated in bold. The genomic region adjacent to MSI2 is included to show genes not bound by MLLT3 (that is, *PCPT* and *VCL*).



Extended Data Fig. 9 | Effects of MLLT3 knockdown and overexpression on CB-HSPC expansion in vitro. a, FACS analysis, 5 and 15 days after MLLT3-KD in CB-HSPCs (representative of two experiments). **b**, Quantification of haematopoietic subsets CD34⁺CD38^{-low} and CD34⁺CD38^{-low}CD90⁺ (HSPCs) in empty vector (CTR) and MLLT3-KD-transduced cells after 5, 15 and 30 days in

culture (mean, $n = 2$). Circles represent individual values. **c**, FACS analysis of CB-HSPCs transduced with control or MLLT3-OE vectors cultured on OP9M2 (representative of two experiments). **d**, Expansion of total live cells (left) and CD34⁺CD38^{-low}CD90 HSPCs (right) transduced with MLLT3-OE or control vectors.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Comparison of the function of MLLT3-OE cells and uncultured cells in transplanted mice. **a**, FACS analysis of total human haematopoietic reconstitution in peripheral blood from mice transplanted with uncultured CB-HSPCs (CD34⁺CD38⁺CD90⁺) or CB-HSPCs transduced with empty GFP vector (CTR) or MLLT3-OE vector after 15 days of culture, as in Fig. 4f (representative of three experiments). Plots show human CD45 haematopoietic reconstitution and multilineage differentiation (myeloid CD66, CD14; B-lymphoid CD19; T-lymphoid CD3, CD4, CD8), and GFP in hCD45⁺ cells. **b**, Quantification of monocytes (CD14⁺), granulocytes (CD66b⁺), B cells (CD14⁺CD66b⁺CD19⁺) and T cells (CD14⁺CD66b⁺CD4⁺ and/or CD8⁺) as a percentage of human CD45⁺ cells in engrafted NSG mice 24 weeks after transplantation (mean, $n = 3$ experiments). P values determined by two-sided t -test. Quantification of human haematopoietic reconstitution in the blood of NSG mice 24 weeks after transplantation in limiting dilution, as in Fig. 4f.

Percentage of hCD45⁺ cells in peripheral blood (**c**) and spleen (**d**) (mean, $n = 3$ experiments). **e**, **f**, FACS plots (**e**) and quantification (**f**) of total human CD45⁺ haematopoietic reconstitution after 18 weeks of secondary transplantation of cord blood (uncultured, day-15-expanded control and MLLT3-OE HSPCs) in mouse bone marrow (mean, $n = 2$ experiments). **g**, **h**, Quantification of human reconstitution in engrafted mice after 24 weeks of primary transplantation. A panel of markers covering the classical AML blast (CD34 and CD33) and B-cell precursor markers (CD34, CD10, CD20, CD24, CD38) were tested in day 15-expanded MLLT3-OE CB-HSPCs, comparing their distribution to mice engrafted with uncultured CB-HSPCs ($n = 5$ mice from two independent experiments, mean and individual values). Haematopoietic markers included CD33⁺ myeloid, CD19⁺ lymphoid, and B-cell precursor markers CD34⁺CD10⁺, CD20⁺, CD24⁺ and CD38⁺ within the CD19⁺ population.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

FACS DATA - DIVA v8
qPCR DATA - Lightcycler Software v1.5.1.62
Sequencing data - Illumina HiSeq 2500 or 4000 softwares

Data analysis

Graph plotting and statistical analysis - Graphpad Prism v8
FACS analysis - Flowjo v10
HTS analysis - bowtie2/2.2.9, tophat/2.0.9 or /2.0.14, samtools/1.3.1, bedtools/2.26.0, bamtools/2.1.0, homer/4.7, deeptools/2.5.0.1, macs2/2.1.1, edgeR v3, Morpheus.
Limiting Dilution Analysis - ELDA

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that support the findings of this study have been deposited in GEO with the accession codes GSE111484. All other data are either available within the paper or from the corresponding authors upon reasonable request. There is no restriction in data availability.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For each experiment, a sample size was chosen to obtain sufficient number of experiments and samples (different human donor tissues) to calculate statistical significance.
Data exclusions	No collected data were excluded from the analysis.
Replication	Data presented were replicated in several independent experiments and by different investigators
Randomization	Allocation to samples or mice to the experimental groups was random.
Blinding	Investigators were not blinded. Data collection for all experiments was automated (e.g. flow cytometry, sequencing etc.) and data interpretation was based on appropriate controls rather than subjective assessment by investigators.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines	<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Antibodies

Antibodies used

human-CD45-BV711 or -BV785 cl.HI30 (304050, 304048; Biolegend, 1:100), mouse-CD45-APC-H7 cl. 30-F11 (557659; BD, 1:100), CD34-APC cl. 581 (555824; BD, 1:20) or -BV605 (343529; Biolegend, 1:20), CD90-FITC cl. 5E10 (555595; BD, 1:100) or -APC (555595; BD, 1:100), CD38-PE-Cy7 or -BUV496cl. HIT2 (560677 and 564657; BD, 1:100), CD19-PE or BV605 cl. 1D3 or HIB19 (12-0193, 12-0199; eBiosciences, 1:50), GPI-80-PE cl. 3H9 (D087-5; MBL, used at 1:50), CD3-PE-Cy7 cl. SK7 (557851, BD) (eBiosciences, 1:50), CD4-APC (MHCD0405; Invitrogen, 1:50), CD8-PE cl. HIT8A (555635; BD, 1:50), CD13-APC cl. WM15 (557454; BD, 1:50), CD66b-BV421 cl. G10F5 (555724; BD, 1:50), CD14-V500 cl. WM53 (561816; BD, 1:100), CD235a-PE or -APC (Glycophorin A, HIR2, BD 1:100), CD71-AF647 (DF1513, Santa Cruz 1:100), CD41a-APC cl. HIP8 (579777; BD, 1:20), CD42b-PEcy7 cl. HIP1 (303916; Biolegend, 1:100), AnnexinV-PE (556422; BD, 1:50), BrdU-PE (556029; BD, 1:100), CD10-AF700 cl. HI10a (563509; BD, 1:100), CD33-PE cl. WM53 (555450; BD, 1:50), CD24-BV711 cl. ML5 (563401; BD, 1:100), CD20-BV650 cl. 2H7 (563780; BD, 1:100), anti-MLLT3, Genetex, GTX102835; anti-V5, Abcam, ab15828; H3K4me3, Abcam ab8580; H3K9ac, CST #9649; H3K9cr, PTM #PTM-516; H3K36me3, Abcam ab9050; H3K79me2, Abcam, ab3594; RNA-polII Rbp1 (8WG16) Biolegend 664911, H3K27ac, AM #39133; H3K4me1, AM #39297; H3K27me3, AM #39155; H3K9m3, AM #39161. anti-CD34 microbeads (Miltenyi Biotec 130-100-453)

Validation

MLLT3 antibody was validated in ChIP by using negative (K562, not expressing MLLT3) and positive (MLLT3-OE, overexpressing MLLT3) cells, as shown in the manuscript. We provide the catalog number, manufacturer and clone (where applicable) for all 40 FACS and ChIP antibodies listed above. All the available information on antibody validation can be found by searching the manufacturer's websites.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	K562, HepG2, 293T (from ATCC). OP9M2, subcloned in our lab from OP9 (see Magnusson, et al Plos One 2013)
Authentication	OP9M2, was extensively authenticated (Magnusson, et al Plos One 2013). The rest of the lines, used only for knockdown validation (K562, HepG2) and virus production (293T), were not validated.
Mycoplasma contamination	Not Tested
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	For human HSC transplantation, 8-12 weeks females from the NSG (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ) strain were used
Wild animals	No wild animals were used in these studies
Field-collected samples	No field collected samples were used in these studies
Ethics oversight	All experiments were carried out in accordance with ethical care guidelines set by University of California on Laboratory Animal Care. Specific protocol numbers available on request.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

GSE111484 is the GEO reference Series
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111484>

Files in database submission

```

INPUT_FUG_ex1_sample12.fastq.gz
INPUT_FUG_ex2_sample17.fastq.gz
INPUT_FUG_ex3_sample19.fastq.gz
INPUT_MLL_ex1_sample13.fastq.gz
INPUT_MLL_ex2_sample18.fastq.gz
INPUT_MLL_ex3_sample20.fastq.gz
K79me2_FUG_ex1_sample1.fastq.gz
K79me2_FUG_ex2_sample1.fastq.gz
K79me2_FUG_ex3_sample5.fastq.gz
K79me2_MLL_ex1_sample5.fastq.gz
K79me2_MLL_ex2_sample9.fastq.gz
K79me2_MLL_ex3_sample13.fastq.gz
V5tag_FUG4w_ex2_sample9.fastq.gz
V5tag_MLL4w_ex2_sample10.fastq.gz
MLLT3_FUG4w_ex1_sample2.fastq.gz
MLLT3_MLL4w_ex1_sample6.fastq.gz
K36m3_FUG4w_ex1_sample4.fastq.gz
K36m3_MLL4w_ex1_sample8.fastq.gz
K4me3_FUG4w_ex1_sample3.fastq.gz
K4me3_MLL4w_ex1_sample7.fastq.gz
H3K4me3_FL_HSPC.fastq.gz
H3K79me2_FL_HSPC_1_sample7.fastq.gz
H3K9ac_FL_HSPC_1_sample14.fastq.gz
INPUT_FL_HSPC_1_sample9.fastq.gz
INPUT_FL_HSPC_2_sample13.fastq.gz
MLLT3_FL_Ery_1_sample11.fastq.gz
MLLT3_FL_Ery_2_sample13.fastq.gz
MLLT3_FL_HSPC_1_sample10.fastq.gz
MLLT3_FL_HSPC_2_sample3.fastq.gz
MLLT3_FL_HSPC_3_sample12.fastq.gz
Pol2_FL_HSPC_ex4.fastq.gz
2-FLd0_90_2_S2_R1_001.fastq.gz

```

3-FLd0_90_3_S3_R1_001.fastq.gz
 4-FLd0_90_4_S4_R1_001.fastq.gz
 CTR1_R1.fastq.gz
 CTR2_R1.fastq.gz
 CTR3_R1.fastq.gz
 CTR4_R1.fastq.gz
 13-FL4w_FU_3_S5_R1_001.fastq.gz
 16-FL4w_FU_4_S8_R1_001.fastq.gz
 MLLT3OE1_R1.fastq.gz
 MLLT3OE2_R1.fastq.gz
 MLLT3OE3_R1.fastq.gz
 MLLT3OE4_R1.fastq.gz
 14-FL4w_Mlo_3_S6_R1_001.fastq.gz
 1FL16_90+d0_ATAC.bw
 FL17_90+d0_ATAC.bw
 HSCd0_HSPC_2.bw
 FL16_Eryd0_ATAC.bw
 FL17_Eryd0_ATAC.bw
 FUG4w_HSPC_1.bw
 MLL4w_HSPC_1.bw
 7-FL4w_Mlo_4_S9_R1_001.fastq.gz
 FL16_90+d0_1.fastq
 FL17_90+d0_2.fastq
 HSCd0_HSPC_2_1.fastq
 FL16_Eryd0_1.fastq
 FL17_Eryd0_1.fastq
 FUG4w_HSPC_1_1.fastq
 MLL4w_HSPC_1_1.fastq
 FL16_90+d0_2.fastq
 FL17_90+d0_2.fastq
 HSCd0_HSPC_2_2.fastq
 FL16_Eryd0_2.fastq
 FL17_Eryd0_2.fastq
 FUG4w_HSPC_1_2.fastq
 MLL4w_HSPC_1_2.fastq
 2-FLd0_90_2_S2_R2_001.fastq.gz
 3-FLd0_90_3_S3_R2_001.fastq.gz
 4-FLd0_90_4_S4_R2_001.fastq.gz
 CTR1_R2.fastq.gz
 CTR2_R2.fastq.gz
 CTR3_R2.fastq.gz
 CTR4_R2.fastq.gz
 13-FL4w_FU_3_S5_R2_001.fastq.gz
 16-FL4w_FU_4_S8_R2_001.fastq.gz
 MLLT3OE1_R2.fastq.gz
 MLLT3OE2_R2.fastq.gz
 MLLT3OE3_R2.fastq.gz
 MLLT3OE4_R2.fastq.gz
 14-FL4w_Mlo_3_S6_R2_001.fastq.gz
 17-FL4w_Mlo_4_S9_R2_001.fastq.gz
 INPUT_FUG4_ex1.bw
 INPUT_FUG_ex2.bw
 INPUT_FUG_ex3.bw
 INPUT_MLL4w_ex1.bw
 INPUT_MLL_ex2.bw
 INPUT_MLL_ex3.bw
 K36me3_FUG4w_ex1.bw
 K36me3_MLL4w_ex1.bw
 K4me3_FUG4w_ex1.bw
 K4me3_MLL4w_ex1.bw
 K79me2_FUG4_ex1.bw
 K79me2_FUG_ex2.bw
 K79me2_FUG_ex3.bw
 K79me2_MLL4_ex1.bw
 K79me2_MLL_ex2.bw
 K79me2_MLL_ex3.bw
 MLLT3_FUG4w_ex1.bw
 MLLT3_MLL4w_ex1.bw
 V5tag_FUG4w_ex2.bw
 V5tag_MLL4w_ex2.bw
 H3K4me3_FL_HSPC.bw
 H3K79me2_FL_HSPC.bw
 H3K9ac_FL_HSPC.bw
 INPUT_FL_HSPC_1.bw
 INPUT_FL_HSPC_2.bw

MLLT3_FL_Ery_1.bw
 MLLT3_FL_Ery_2.bw
 MLLT3_FL_HSPC_1.bw
 MLLT3_FL_HSPC_2.bw
 MLLT3_FL_HSPC_3.bw
 RNApolIII_FL_HSPC.bw
 diffOE4w_6s_paired_edgeR.txt
 MLLT3OE4w_6s_d0_3s_exoRpkMSt.txt

Genome browser session
 (e.g. [UCSC](#))

No Longer applicable

Methodology

Replicates

All ChIPs performed to study the MLLT3-OE effect on FL-HSPC were repeated at least 2 times: (ChIPseq K79 MLLT3OE vs CTR n=5). Some genomic distribution data of histone marks on uncultured HSPC (Fig.3 c,e) were performed only once and checked against deposited data on similar cell types.

Sequencing depth

RNAseq libraries were multiplexed in 4-5 samples per lane of a Nextseq run in PE75
 ChIPseq libraries were multiplexed in 8-10 samples per lane of a Hiseq4000 run in SE50
 ATACseq libraries were multiplexed in 4-6 samples per lane of a Hiseq2500 run in PE50
 Sequencing depth for samples in Fig. 5 expressed as uniquely aligned reads (or pairs for RNAseq) to hg19 genome:

Assay Sample Exp number UMR
 RNAseq Uncult. ex1 17393805
 RNAseq Uncult. ex2 19671545
 RNAseq Uncult. ex3 15297822
 RNAseq CTR ex1 16996453
 RNAseq CTR ex2 16227199
 RNAseq CTR ex3 16653389
 RNAseq CTR ex4 14875982
 RNAseq CTR ex5 16022515
 RNAseq CTR ex6 15218252
 RNAseq MLLT3-OE ex1 16798850
 RNAseq MLLT3-OE ex2 17009253
 RNAseq MLLT3-OE ex3 16186420
 RNAseq MLLT3-OE ex4 16155541
 RNAseq MLLT3-OE ex5 15010228
 RNAseq MLLT3-OE ex6 14440138
 K79me2_IP CTR ex1 10615366
 K79me2_IP CTR ex2 23718613
 K79me2_IP CTR ex3 67495255
 K79me2_IP CTR ex4 16373101
 K79me2_IP CTR ex5 15993267
 K79me2_IP MLLT3-OE ex1 7440663
 K79me2_IP MLLT3-OE ex2 44784341
 K79me2_IP MLLT3-OE ex3 59145909
 K79me2_IP MLLT3-OE ex4 11596427
 K79me2_IP MLLT3-OE ex5 48362880
 K4me3_IP CTR ex1 13264734
 K4me3_IP CTR ex2 14809389
 K4me3_IP MLLT3-OE ex1 12465814
 K4me3_IP MLLT3-OE ex2 12946395
 K36me3_IP CTR ex1 92828637
 K36me3_IP CTR ex2 13751739
 K36me3_IP MLLT3-OE ex1 55130829
 K36me3_IP MLLT3-OE ex2 12898815
 K9ac_IP CTR ex1 16407526
 K9ac_IP CTR ex2 17368673
 K9ac_IP MLLT3-OE ex1 30788024
 K9ac_IP MLLT3-OE ex2 9354776
 K79me2_IP CTR-KD ex1 24878114
 Antibodies
 Peak calling parameters
 Data quality
 Software
 Flow Cytometry
 Plots
 Confirm that:
 K79me2_IP CTR-KD ex2 57520064
 K79me2_IP MLL T3-KD ex1 63125697
 K79me2_IP MLL T3-KD ex2 27133294
 K79me2_IP CTR-OE d10 ex1 17693670
 K79me2_IP CTR-OE d10+EPZ5676 ex1 23559122

	K79me2_IP MLLT3-OE d10 exl 33609548 K79me2_IP MLLT3-OE d10+EPZ5676 exl 15271768
Antibodies	anti-MLLT3, Genetex, GTX102835; anti-V5, Abcam, ab15828; H3K4me3, Abcam ab8580; H3K9ac, CST #9649; H3K9cr, PTM #PTM-516; H3K36me3, Abcam ab9050; H3K79me2, Abcam, ab3594; RNA-poli I Rbpl (8WG16) Biolegend 664911, H3K27ac, Active Motif #39133; H3K4me1, AM #39297; H3K27me3, AM #39155; H3K9m3, AM #39161
Peak calling parameters	MLLT3 Peaks were called with macs2, using broad Peak algorithm and default parameters
Data quality	Quality of the reads was checked with the fastqc software. Quality of the alignment was checked with samtools flagstat.
Software	HTS analysis - bowtie2/2.2.9, tophat/2.0.9 or /2.0.14, samtools/1.3.1, bedtools/2.26.0, bamtools/2.1.0, homer/4.7, deeptools/2.5.0.1, macs2/2.1.1,

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	FACS analysis was performed using single cell suspensions prepared from human fetal liver and cord blood after isolation from the tissues or culture, or obtained from NSG mice transplanted with human cells. Cells were stained in PBS 5%FBS.
Instrument	Cells were assayed on a BD-LSRII flow cytometer. Cell sorting was performed using a BD FACS Aria II.
Software	BD DIVA v8 FlowJo software (Tree Star Inc.)
Cell population abundance	Cell purity was checked periodically when setting up the sorting instruments, resulting in >95 % purity
Gating strategy	For both flow cytometry analysis and sorting, cells were selected for live cell scatter in FSC/SSC then for singlets in FSCa/FSCb or FSCb/FSCw and SSCb/SSCw then for 7AAD negativity (dead cell exclusion). Other markers were determined positive when signal was above FMO (fluorescence minus one) control.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Patch repair of deep wounds by mobilized fascia

<https://doi.org/10.1038/s41586-019-1794-y>

Received: 28 February 2019

Accepted: 30 October 2019

Published online: 27 November 2019

Donovan Correa-Gallegos^{1,7}, Dongsheng Jiang^{1,7}, Simon Christ¹, Pushkar Ramesh¹, Haifeng Ye¹, Juliane Wannemacher¹, Shruthi Kalgudde Gopal¹, Qing Yu¹, Michaela Aichler², Axel Walch², Ursula Mirastschijski^{3,4}, Thomas Volz⁵ & Yuval Rinkevich^{1,6*}

Mammals form scars to quickly seal wounds and ensure survival by an incompletely understood mechanism^{1–5}. Here we show that skin scars originate from prefabricated matrix in the subcutaneous fascia. Fate mapping and live imaging revealed that fascia fibroblasts rise to the skin surface after wounding, dragging their surrounding extracellular jelly-like matrix, including embedded blood vessels, macrophages and peripheral nerves, to form the provisional matrix. Genetic ablation of fascia fibroblasts prevented matrix from homing into wounds and resulted in defective scars, whereas placing an impermeable film beneath the skin—preventing fascia fibroblasts from migrating upwards—led to chronic open wounds. Thus, fascia contains a specialized prefabricated kit of sentry fibroblasts, embedded within a movable sealant, that preassemble together diverse cell types and matrix components needed to heal wounds. Our findings suggest that chronic and excessive skin wounds may be attributed to the mobility of the fascia matrix.

Mammalian scarring occurs when specialized fibroblasts migrate into wounds to deposit plugs of extracellular matrix¹. Abnormal scarring results in either non-healing chronic wounds or aggravating fibrosis^{2–4}, which represent a high burden for patients and to the global healthcare system. In the US alone, costs related to impaired scarring amount to tens of billions of dollars per year⁵.

The origin of fibroblasts in wounds and the mechanism by which they act remain unclear⁶. Possible sources, such as papillary (upper) and reticular (lower) dermal layers⁷, pericytes⁸, adipocytes^{9,10} and bone-marrow derived monocytes¹¹ have been suggested. We previously demonstrated that all scars in the back skin derive from a distinct fibroblast lineage expressing the engrailed 1 gene (*En1*) during embryogenesis and we refer to these cells as *En1*-lineage positive fibroblasts (EPFs)^{12,13}. This lineage is present not only in the skin but also in the strata underneath, called fascia.

The fascia is a gelatinous viscoelastic membranous sheet that creates a frictionless gliding interface between the skin and the body's rigid structure below. Mouse back-skin fascia extends as a single sheet separated from the skin by the Panniculus carnosus (PC) muscle, whereas in humans back-skin there is no intervening muscle and the fascia consists of several thicker sheets that are continuous with the skin. In humans the fascia layers incorporate fibroblasts, lymphatics, adipose tissue, neurovascular sheets and sensory neurons^{14,15}.

Here we explored the fundamental mechanisms of scar formation by using matrix-tracing techniques, live imaging, genetic-lineage tracing and anatomic fate-mapping models. We identified the fascia as a major source of wound-native cells, including fibroblasts. Notably, we found that wound provisional matrix originates from prefabricated matrix in the fascia that homes in to open wounds as a movable sealant dragging along vasculature, immune cells and nerves, upwards into the skin.

Wound cells rise from fascia

To trace the origins of cells in wounds, we developed a fate-mapping technique by transplanting chimeric skin and fascia grafts into living animals (Fig. 1a and Methods).

At 14 days post-wounding (dpw), $80.04 \pm 3.443\%$ (mean \pm s.e.m.) of the labelled cells in the wound originated from fascia (Fig. 1b). Fascia-derived cells filled the entire wound bordering the regenerated epidermis and even the surrounding dermis, making up $35.46 \pm 4.938\%$ of the total labelled cells within a 0.2 mm radius (Fig. 1b, c). ACTA2⁺ myofibroblasts ($81.63 \pm 12.84\%$), nerve cells, endothelial cells and macrophages within wounds were predominantly of fascia origin (Fig. 1d, e). Independently, in vivo labelling of the fascia showed same results (Extended Data Fig. 1a, Methods). Labelled cells populated the wounds and surrounding dermis at 14 dpw, whereas in uninjured controls, labelled cells remained in the fascia (Extended Data Fig. 1b). Up to $56.71 \pm 9.319\%$ of fascia-derived cells in wounds expressed classical fibroblast markers (Extended Data Fig. 1c). Labelled monocytes (macrophages), lymphatics, endothelium and nerves also derived from fascia (Extended Data Fig. 1d). Collectively, our two independent fate-mapping approaches demonstrate that fascia is a major reservoir of fibroblasts, endothelial cells macrophages and peripheral nerves that populate wounds following injury.

Fascia fibroblasts dictate scar severity

We then analysed the scar-forming EPFs across dermal and fascia compartments by using a TdTomato-to-GFP replacement reporter^{12,13} (*En1^{cre};R26^{mTmG}*; see Methods). Fibroblasts were the predominant fascia cell type (71.1%), whereas dermis had a significantly lower fraction of

¹Group Regenerative Biology and Medicine, Institute of Lung Biology and Disease, Helmholtz Zentrum München, Munich, Germany. ²Research Unit Analytical Pathology, Helmholtz Zentrum München, Munich, Germany. ³Mira-Beau Gender Esthetics, Berlin, Germany. ⁴Wound Repair Unit, CBIB, Department of Biology and Biochemistry, University of Bremen, Bremen, Germany. ⁵Department of Dermatology and Allergology, School of Medicine, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany. ⁶German Centre for Lung Research (DZL), Munich, Germany. ⁷These authors contributed equally: Donovan Correa-Gallegos, Dongsheng Jiang. *e-mail: yuval.rinkevich@helmholtz-muenchen.de

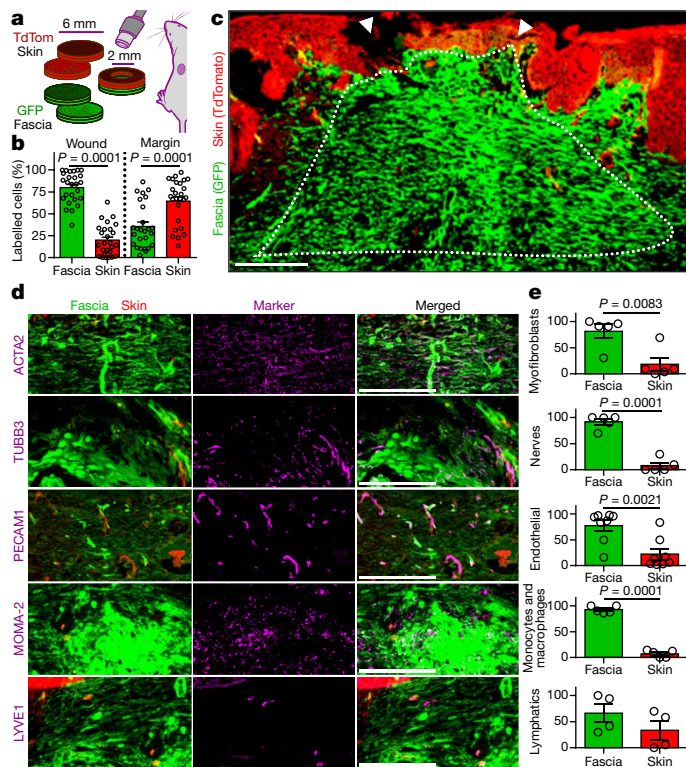


Fig. 1 | Fascia is a major cellular source for wounds. **a**, Generation of 6-mm diameter chimeric grafts with an inner 2-mm wound to determine contributions of dermis and fascia. **b**, Percentages of TdTomato⁺ or GFP⁺ cells from total labelled cells in the wound and wound margin. Data are mean \pm s.e.m., $n = 26$ images from 4 biological replicates. One-way ANOVA, multiple comparison Tukey's test, 95% confidence interval (CI). **c**, Wound showing skin- and fascia-derived cells at 14 dpw. **d**, **e**, Immunolabelling and contributions of myofibroblasts (ACTA2), nerves (TUBB3), blood vessels (PECAM1), monocytes or macrophages (MOMA-2) and lymphatic cells (LYVE1). Data are mean \pm s.e.m., $n = 4$ (LYVE1), 5 (ACTA2, TUBB3 and MOMA-2) or 8 (PECAM1) images from 4 biological replicates. Unpaired two-tailed t -test, 95% CI. Dotted lines delimit the wound. Arrowheads indicate injury site. Scale bars, 200 μ m. TdTom, TdTomato.

fibroblasts (56.4%, Extended Data Fig. 2a, b). Within this population, there were twice as many EPFs as En1-naïve fibroblasts (ENFs) in the fascia (61.2% and 31.8%, respectively). In dermis, there was a sixfold excess of EPFs (83.13% EPFs versus 12.78% ENFs; Extended Data Fig. 2c, d). Fascia was also enriched in regenerative cell types such as endothelial cells and lymphatics, whereas populations of macrophages and nerve cells were similar in both compartments (Extended Data Fig. 2e). Thus, a higher proportion of fibroblasts, endothelial cells and lymphatic cells and a lower EPF to ENF ratio distinguishes the fascia from dermis.

Two-photon microscopy revealed that fascia EPFs assemble in monolayers of consecutive perpendicular sheets across the dorsal–ventral axis (Extended Data Fig. 2f, Supplementary Video 1). EPFs populate the entire back in topographic continua extending from the fascia and traversing the PC (Extended Data Fig. 2g, h Supplementary Video 2). Regions where PC ended or where nerve bundles and blood vessels traversed the PC also showed continuums of EPFs without clear boundaries (Extended Data Fig. 2i, j). To test whether fascia EPFs could access dermal layers upon injury, we generated superficial excisional wounds. Aggregates of EPFs rising into open wounds from fissures in the PC were observed after only 3 dpw (Extended Data Fig. 2k, Supplementary Video 3). Collectively, our observations suggest that fascia EPFs easily traverse upwards into dermal layers during wounding and are unobstructed by the PC muscle.

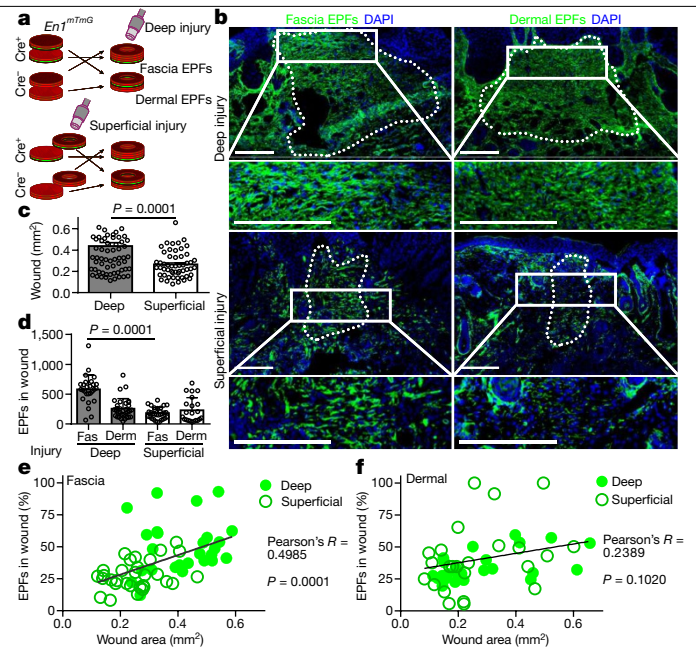


Fig. 2 | Fascia EPFs dictate scar severity. **a**, Dermal versus fascia EPFs chimeric grafts. Chimeric grafts were prepared as in Fig. 1a, using the EPF reporter line *En1*^{Cre}; *R26*^{mTmG} combined with Cre-negative littermate samples to trace EPFs in each compartment. **b**, Images showing fascia EPFs (left) or dermal EPFs (right) in wounds after a deep (top) or superficial injury (bottom). **c**, Wound size. Data are mean \pm s.e.m., $n = 70$ (deep) and 53 (superficial) images analysed from 5 biological replicates. Unpaired, two-tailed t -test, 95% CI. **d**, Numbers of fascia and dermal EPFs. Data are mean \pm s.e.m., $n = 27, 32, 27$ and 22 images analysed from 5 biological replicates. Unpaired, two-tailed t -test, 95% CI. **e**, **f**, Plots of EPF fractions and wound size from fascia (**e**) and dermal EPFs (**f**). $n = 57$ (**e**) and 48 (**f**) images analysed from 5 biological replicates. Two-tailed Pearson's R^2 correlation, 95% CI. Dotted lines delimit wounds. Scale bars, 200 μ m.

The size of a scar increases with the depth of the injury¹⁶. We therefore investigated whether this correlation can be attributed to fascia by analysing the extent of fibroblast contributions from the fascia and dermis in deep versus superficial wounds. For this, we combined genetic lineage-tracing (*En1*^{Cre}; *R26*^{mTmG}) with our anatomic fate-mapping chimeric grafts and inflicted superficial or deep injuries (Fig. 2a, Methods). Fourteen days post-wounding, mean wound size of deep injuries was 1.7 times that of superficial injuries (Fig. 2b, c). There were twice as many fascia EPFs in deep wounds, whereas the number of dermal EPFs remained constant in both conditions (Fig. 2d). The abundance of fascia EPFs in the wound directly correlated with wound size and thus scar severity, whereas dermal EPFs showed no such correlation (Fig. 2e, f). No crossing of EPFs between these compartments was observed in uninjured controls, indicating that the influx of fascia EPFs was triggered by injury (Extended Data Fig. 3a, b).

Long-term tracing of fascia EPFs in wounds showed that they recede ten weeks after injury (Extended Data Fig. 3c). This desertion from mature scars occurred through an apoptosis-independent mechanism, indicated by a low rate of cell death (<5%) across earlier time points (Extended Data Fig. 3d, e).

We next sought to place fascia EPFs in the framework of known lineage markers used to define different populations of wound fibroblasts such as CD24, CD34, DPP4, DLK1 and LY6A^{7,10,12}. All markers were prominent in fascia EPFs and were surprisingly downregulated upon entering the wound in our graft experiments (Extended Data Fig. 4). Flow cytometry confirmed the higher expression of DPP4, ITGB1, LY6A and PDGFR α in fascia than in dermal fibroblasts (Extended Data Fig. 5a–c). Sorted fascia EPFs also revealed low cellular heterogeneity, with the population predominantly comprising LY6G⁺PDGFR α ⁺ (87.0%) and DPP4⁺ITGB1⁺

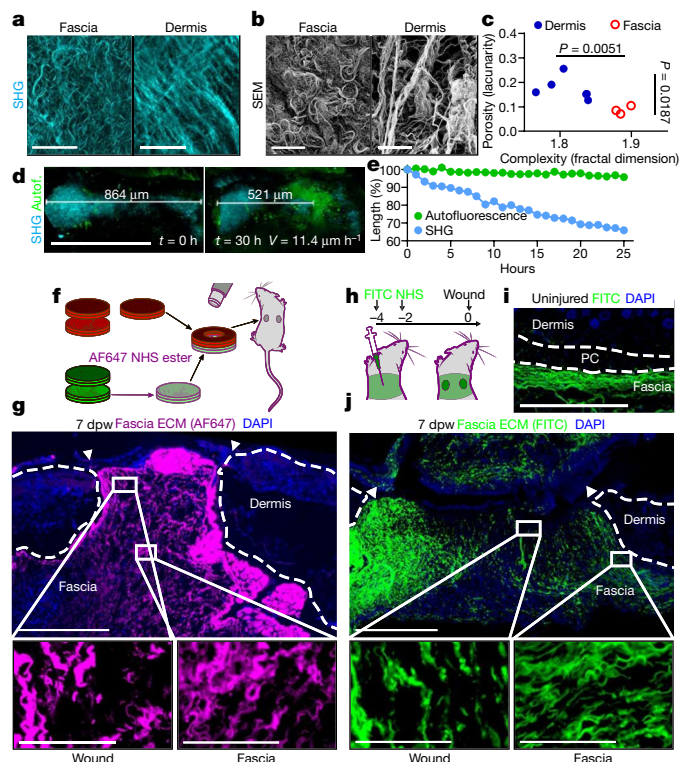


Fig. 3 | Fascia matrix steers into wounds. **a**, SHG (**a**) and SEM images (**b**) of fascia (left) and dermis (right) showing matrix arrangements, representative images of 3 biological replicates. **c**, Fractal dimension and lacunarity values from SHG images. $n = 6$ (dermis) and 3 (fascia) images analysed from 3 biological replicates. Unpaired two-tailed t -test, 95% CI. **d**, Time-lapse images of fascia in culture at time 0 (left) and 30 h (right). Representative video from at least three independent experiments. Autof., autofluorescence. **e**, Contraction rate from the SHG and autofluorescence. Data derived from Supplementary Video 4. **f**, Fascia matrix labelling with AF647 NHS ester in chimeric grafts. TdTomato and GFP reporter lines biopsies were used for the chimeric grafts as before. Only the matrix of the GFP-positive fascia compartment was labelled before transplantation. **g**, Images at 7 dpw showing fascia matrix covering the wound. Representative image of at least three biological replicates. **h**, In situ fascia matrix-tracing experiment using FITC NHS ester. Subcutaneous injections of FITC NHS ester were performed 4 and 2 days before wounding in the back-skin of wild-type mice to label the fascia matrix. **i**, Uninjured controls showing specificity of the labelling in fascia. Representative sample of at least three biological replicates. **j**, Images from 7 dpw showing fascia matrix covering the wound area. Representative samples of at least three biological replicates. Arrowheads indicate the original injury. Broken lines delimit dermis (**g** and **j**) or PC (**i**). Scale bars, 30 μm (**a**, **b**), 500 μm (**d**, **g**, **i**, **j**), and 100 μm (**g**, **j** magnified insets).

(72.8%) cells (Extended Data Fig. 5d). This broad marker convergence identifies fascia EPFs as the major source of wound fibroblasts.

Provisional matrix emerges from fascia matrix

We then looked at the fascia matrix itself. Second harmonic generation (SHG) signal and scanning electron micrographs (SEM) revealed profuse coiled collagen fibrils in the fascia, indicative of a relaxed and immature matrix (Fig. 3a, b). Fractal measurements¹³ of the fibre alignments showed a more condensed matrix configuration in fascia than the stretched and woven dermal matrix (Fig. 3c).

Given the immaturity of the fascia matrix, we checked whether it could work as a repository for scar tissue. We developed an incubation chamber that enabled live imaging of fascia biopsies over several days (Methods). Recording of SHG signals showed steering of the matrix at a rate of $11.4 \mu\text{m h}^{-1}$ (Fig. 3d, e, Supplementary Video 4). Assuming

a similar rate in vivo, the fascia matrix could move about 2 mm in 7 days, accounting for the dynamics of provisional matrix deposition in mammals.

To test whether fascia matrix steers into wounds in vivo, we developed a method to trace the fascia matrix in our chimeric grafts using NHS esters. (Fig. 3f, Methods). Streams of traced matrix from the fascia extended upwards and plugged the open wounds from 7 dpw (Fig. 3g, Extended Data Fig. 6a, b). Fascia-derived matrix covered $74.78 \pm 12.94\%$ of total collagen content in the wound (Extended Data Fig. 6c). Our data indicated that individual fascia matrix fibres were not being pulled; instead, pliable matrix was extended upwards to mould the wound. Advanced wound stages showed a decline in label in specific regions of the wound, suggesting an active remodelling process of the fascia-derived matrix (Extended Data Fig. 6d–f, Supplementary Video 5).

We then tested whether dermal matrix could be steered by double labelling our chimeric grafts. Only the fascia matrix plugged deep injury wounds (Extended Data Fig. 6g–j), whereas dermal matrix remained immobile in deep and superficial injuries; the superficial injuries healed via de novo matrix deposition (Extended Data Fig. 6k, l).

To provide further evidence showing that fascia matrix migrates to open wounds, we labelled the fascia matrix in situ before injury (Fig. 3h, i, Methods). Labelled matrix made up most of the wound provisional matrix (Fig. 3j), which underwent remodelling during the first two weeks following injury (Extended Data Fig. 7a, b). Fractal measurements showed that fascia fibre interfaces expanded by 3 dpw, changing from a parallel sheet arrangement to a highly porous plug. This expansion was followed by contraction into thicker and more complex mature scar matrix architecture (Extended Data Fig. 7c–e). Surprisingly, traced matrix was also present in the eschar. Activated platelets infiltrated and clustered within fascia fibres before eschar formation (Extended Data Fig. 7f), indicating that the coagulation cascade occurs in parallel with fascia matrix steering.

EPFs steer fascia matrix into wounds

To test whether matrix steering from fascia is caused by EPFs, we blocked fascia by implanting an impermeable dual surface expanded polytetrafluoroethylene (ePTFE) membrane¹⁷ between the fascia and the PC in wounds of *Enl^{cre};R26^{VT2/GK3}* mice (Fig. 4a, Methods). Surprisingly, wounds with implants remained completely open whereas sham controls closed within 21 days (Fig. 4b). After two months, EPFs trailed from the wound margins and under the membrane without generating scars (Fig. 4c, Extended Data Fig. 8a). Implants produced a transient inflammation that resolved even when the wound remained open. Two-month-old wounds with implants showed normal leukocyte and pro- and anti-inflammatory interleukins levels (Extended Data Fig. 8b–i), consistent with the clinical use of ePTFE as immunologically inert membranes. The coagulation cascade also was unaltered at the border between the dermis and the membrane (Extended Data Fig. 8j, k). These results indicate that the lack of scarring with ePTFE membranes does not reflect chronic inflammation or poor clotting, but rather a blockade of fascia steering mediated by the fascia fibroblasts. These findings further support the notion that scar tissue is mostly derived from the fascia, since dermal EPFs or dermal matrix are unable to repair wounds in the absence of fascia movements.

We next investigated whether mechanical separation between dermis and fascia alone, without barrier implants, would affect matrix steering and scar formation. To address this question, we performed full excisional wounds in wild-type mice and physically released the fascia below the PC around the wound (Fig. 4d). Wound closure from released-fascia wounds was significantly delayed, and wounds remained open early on, similarly to those documented following membrane implantations (Fig. 4e, f).

To definitively link fascia EPFs to matrix steering, we genetically ablated fascia EPFs using two separate strategies. First, we used a

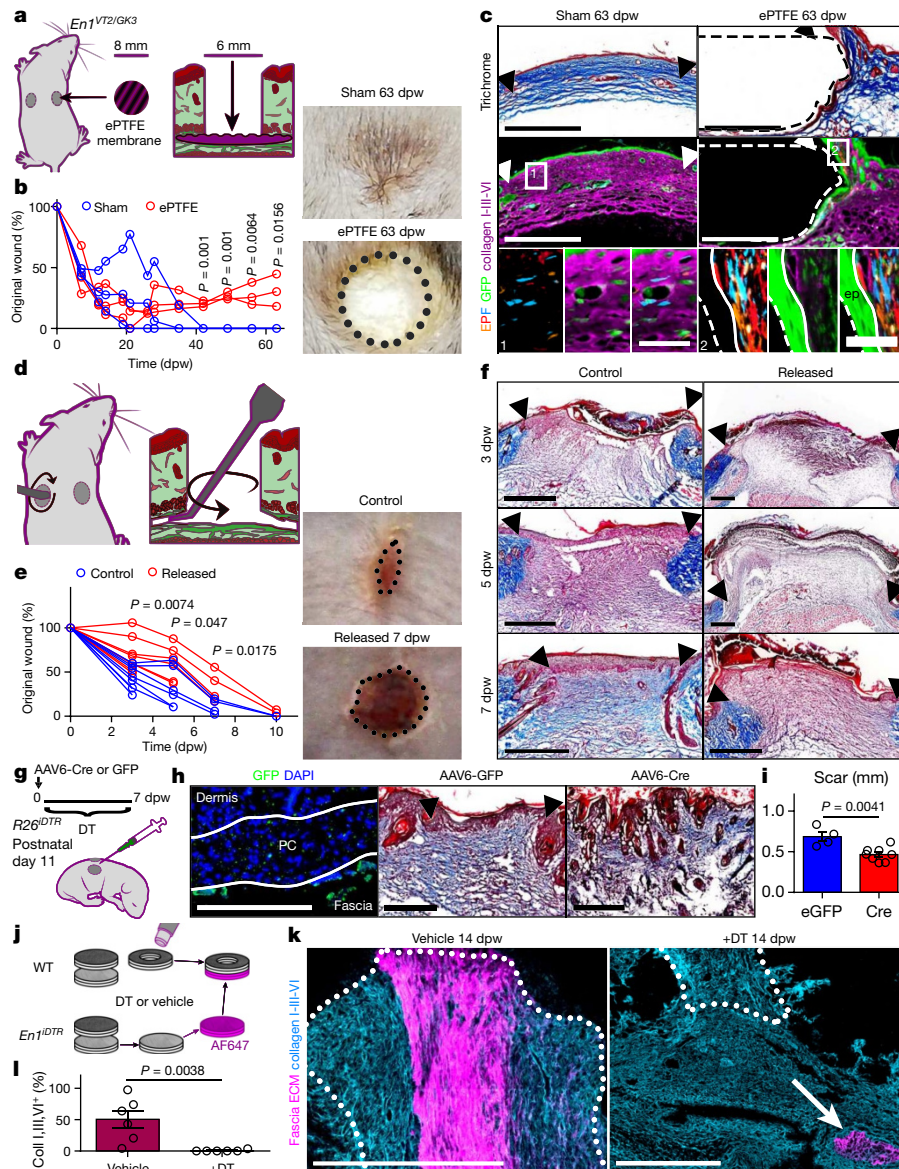


Fig. 4 | Fascia EPFs steer scar primordium into wounds. **a**, 8-mm diameter ePTFE membranes were implanted between the skin and fascia in fresh 6-mm diameter wounds to block fascia steering. **b**, ePTFE-implanted or sham wound closure (left) determined from photographs (right) at specified time points. Individual values, $n = 3$ biological replicates. Unpaired two-tailed t -test, 95% CI. **c**, Sham (left) or ePTFE-implanted (right) wounds at 63 dpw. Trichrome staining (top) and collagen immunolabelling (middle). Magnifications 1, 2 (bottom) show multiclonal dermal EPFs. Representative images from 3 biological replicates. **d**, Fascia tissue around fresh 5-mm diameter wounds was mechanically separated from the skin above using a metal paddle. **e**, Fascia-released or control wound closure (left) determined from photographs (right) at specified time points. Individual values, $n = 8$ (0–3 dpw), 6 (5 dpw), 4 (7 dpw) and 2 (10 dpw) images from 8 biological replicates. Unpaired two-tailed t -test, 95% CI. **f**, Trichrome-stained wounds at 3, 5 and 7 dpw from control or fascia-

released wounds. Representative images from 8 biological replicates. **g**, Fascia cell depletion in *R26^{IDTR}* neonates. **h**, Fluorescence image (left) and trichrome-stained wounds at 7 dpw in GFP- (middle) or Cre-transduced (right) fascia. Representative images from 3 biological replicates. **i**, Scar-length measurements. Data are mean \pm s.e.m.; $n = 4$ and 8 sections from 3 biological replicates. Unpaired two-tailed t -test, 95% CI. **j**, Depletion of fascia EPFs in chimeric grafts with fascia matrix labelling. **k**, Immunolabelling for collagens, and fascia matrix in control (left) or DT-treated (right) grafts. Representative images from 3 biological replicates. **l**, Matrix-labelling coverage. Data are mean \pm s.e.m.; $n = 6$ sections from 3 biological replicates. Unpaired two-tailed t -test, 95% CI. Dashed lines delimit the implant. Dotted lines delimit the wound. Arrowheads indicate the original injury. Arrow indicates remaining labelled fascia matrix in DT-treated grafts. Scale bars, 50 μ m (c (1 and 2)), 200 μ m (h), and 500 μ m (c (main image), f, h, k).

transgenic line that expresses the diphtheria toxin receptor (DTR) in a Cre-dependent manner (*R26^{IDTR}*), enabling us to deplete cells expressing Cre recombinase upon exposure to diphtheria toxin (DT). We thus generated Cre-expressing adeno-associated viral particles (AAV6-Cre) and injected them into the fascia of *R26^{IDTR}* pups underneath freshly made full excisional wounds (Fig. 4g). Scar size from AAV6-Cre transduced mice treated with DT were significantly smaller those from controls (Fig. 4h, i).

Second, we used *En1^{cre};R26^{IDTR}* double transgenic mice, in which DTR expression is restricted to EPFs, making them susceptible to

DT-mediated ablation. We corroborated the ablation of fascia EPFs in cultured biopsies 6 days after acute exposure to DT for 1 h. Effective dose of DT prevented the normal increase in collagen fibre density observed in control samples and decreased the cell density by 2.5 times (Extended Data Fig. 9a–c). Live imaging showed absence of any matrix steering after DT exposure (Extended Data Fig. 9d, e, Supplementary Video 6), confirming that fascia EPFs are essential for matrix steering.

Next, we created chimeric grafts using dermis from wild-type mice and fascia from *En1^{cre};R26^{IDTR}* mice. We ablated fascia EPFs using DT

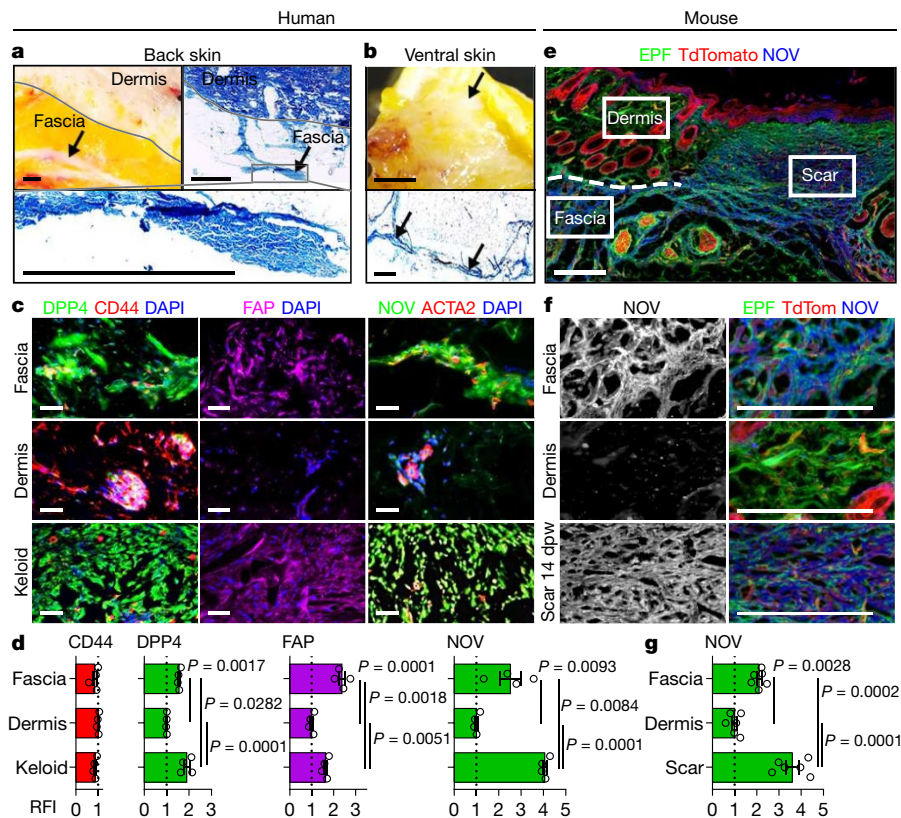


Fig. 5 | Marker signature in keloids and fascia. **a, b**, Macroscopic images and trichrome staining of human back (**a**) and abdominal (**b**) skin showing fascia layers embedded in subcutaneous tissue. Arrows indicate the fascia. Representative images from 4 independent samples. **c**, Immunolabelling for DPP4, CD44 and FAP, and NOV and ACTA2 in fascia, dermis and keloids of human back skin. **d**, Relative fluorescence intensity (RFI) in **c**, normalized to the dermis. Data are mean \pm s.e.m.; $n = 4$ images analysed from 4 biological

replicates. One-way ANOVA with Tukey's test, 95% CI. **e, f**, Immunostaining for NOV in *En1^{cre};R26^{mTg}* 14 dpw scars. **g**, Relative fluorescence intensity in **f**, normalized to the dermis. Data are mean \pm s.e.m.; $n = 6$ images of 3 biological replicates. One-way ANOVA, Tukey's test, 95% CI. Dotted and broken lines delimit scar and fascia, respectively. Scale bars, 2 mm (**a, b**), 50 μ m (**c**) and 200 μ m (**e, f**).

and labelled the matrix before transplantation (Fig. 4j). Ablation of fascia EPFs prevented matrix steering into the wound (Fig. 4k, l) and instead labelled matrix remained in the fascia layer below. Together, our data demonstrate that fascia resident EPFs actively steer matrix to seal open wounds.

To check whether fibroblast proliferation preceded and was needed for matrix steering, we analysed the proliferation rate in our matrix-tracing experiments. Expansion of the fascia gel beneath the wound occurred during the first days after injury, whereas cell proliferation peaked after one week (Extended Data Fig. 10a–c), indicating that proliferation is not required for matrix steering. Furthermore, treatment with a proliferation inhibitor had no effect on fascia matrix steering in vitro (Extended Data Fig. 9f–k, Supplementary Video 7). Our results demonstrate that fascia matrix works as an expanding sealant that quickly clogs deep wounds independently of cell proliferation.

Fascia and keloid share marker signatures

Similar to mouse fascia matrix, human keloids present poorly structured and densely packed collagen fibres¹⁸. This motivated us to investigate the presence of fascia fibroblasts in keloid tissue. We screened for markers present in keloids and compared them with healthy dermis and the connective tissue in the subcutaneous space (fascia) of human skin across multiple anatomic locations (Fig. 5a, b). FAP and DPP4 were highly expressed in both fascia and keloids, with low expression in dermis. The fascia-restricted protein NOV was also prominently expressed in both human and mouse fascia, as well as in human keloids and mouse scars (Fig. 5c–g). This preservation of fascia markers across mouse

and human fascia and keloid scars suggests a common fascia origin for most human cutaneous scars.

Discussion

Current models of wound healing propose that dermal fibroblasts migrate into wounds and locally deposit matrix de novo onto the granulation tissue provided by the coagulation cascade. This provisional matrix is then remodelled into a mature scar. On the basis of our findings in this study, we propose a revised model (Fig. 6) in which, in deep injuries, fascia fibroblasts pilot their local composite matrix into wounds that, in coordination with the coagulation cascade, form the provisional matrix. Thus, instead of de novo matrix deposition by dermal fibroblasts, a 'scar primordium' is steered by the fascia fibroblasts. Thus, fascia serves as an external repository, or *externum repono*, of scar-forming provisional matrix, which represents an efficient mechanism to quickly seal large open wounds. Previous studies have shown that the matrix undergoes movement during early development and organ morphogenesis^{19–22}. To our knowledge, the extent and magnitude of matrix movements that we document here have not been observed during injury or regenerative settings. Cultured dermal fibroblasts have been shown to pull and reorient individual collagen or fibronectin fibres locally in cultured plates and in 3D in vitro assays^{23,24}. However, our findings reveal highly dynamic and large-scale movements of composite tissue matrix during injury that are mediated exclusively by specialized fibroblasts of the fascia.

Our findings on the contribution of fascia to large scars and its blockage leading to chronic open wounds indicates that the spectrum of poor

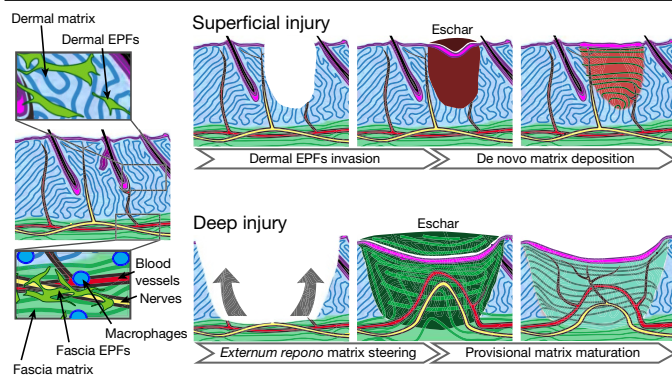


Fig. 6 | Revised wound healing model. Superficial injuries heal by the classical fibroblast migration and de novo matrix deposition process. In response to a deep injury, fascia fibroblasts steer their surrounding tissue into wounds. Fascia-derived macrophages, endothelial and peripheral nerves rapidly clog the open wound. In coordination with the platelet response, the fascia matrix serves as a provisional matrix that undergoes remodelling until it forms a mature scar.

and excessive scarring in the skin, such as in diabetic and ulcerative wounds, as well as in hypertrophic and particularly keloid scars, might all be attributed to fascia. Indeed, the subcutaneous fascia varies widely according to species, sex, age and anatomic skin location²⁵. In some mammals, the superficial fascia is loose, whereas in scar-prone species such as human, dog and horse, the superficial fascia is thicker. Human fascia further varies in thickness in different regions of the body²⁶. For example, the lower chest, back, thigh and arm have much thicker and multi-layered membranous sheets, and it is these anatomic sites that are prone to form hypertrophic and keloid scars²⁷. Understanding the topographic anatomy of the fascia layer may help explain scar phenotypes and severities, including the occurrences of hypertrophic and keloid scars.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1794-y>.

1. Marshall, C. D. et al. Cutaneous scarring: basic science, current treatments, and future directions. *Adv. Wound Care* **7**, 29–45 (2018).

2. Finnerty, C. C. et al. Hypertrophic scarring: the greatest unmet challenge after burn injury. *Lancet* **388**, 1427–1436 (2016).
3. Morton, L. M. & Phillips, T. J. Wound healing and treating wounds: differential diagnosis and evaluation of chronic wounds. *J. Am. Acad. Dermatol.* **74**, 589–605, quiz 605–606 (2016).
4. Do, N. N. & Eming, S. A. Skin fibrosis: models and mechanisms. *Curr. Res. Transl. Med.* **64**, 185–193 (2016).
5. Sen, C. K. et al. Human skin wounds: a major and snowballing threat to public health and the economy. *Wound Repair Regen.* **17**, 763–771 (2009).
6. Hinz, B. Myofibroblasts. *Exp. Eye Res.* **142**, 56–70 (2016).
7. Driskell, R. R. et al. Distinct fibroblast lineages determine dermal architecture in skin development and repair. *Nature* **504**, 277–281 (2013).
8. Greenhalgh, S. N., Conroy, K. P. & Henderson, N. C. Healing scars: targeting pericytes to treat fibrosis. *QJM* **108**, 3–7 (2015).
9. Plikus, M. V. et al. Regeneration of fat cells from myofibroblasts during wound healing. *Science* **355**, 748–752 (2017).
10. Shook, B. A. et al. Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair. *Science* **362**, eaar2971 (2018).
11. Mori, L., Bellini, A., Stacey, M. A., Schmidt, M. & Mattoli, S. Fibrocytes contribute to the myofibroblast population in wounded skin and originate from the bone marrow. *Exp. Cell Res.* **304**, 81–90 (2005).
12. Rinkevich, Y. et al. Identification and isolation of a dermal lineage with intrinsic fibrogenic potential. *Science* **348**, aaa2151 (2015).
13. Jiang, D. et al. Two succeeding fibroblastic lineages drive dermal development and the transition from regeneration to scarring. *Nat. Cell Biol.* **20**, 422–431 (2018).
14. Adstrum, S., Hedley, G., Schleip, R., Stecco, C. & Yucesoy, C. A. Defining the fascial system. *J. Bodyw. Mov. Ther.* **21**, 173–177 (2017).
15. Stecco, C. & Schleip, R. A fascia and the fascial system. *J. Bodyw. Mov. Ther.* **20**, 139–140 (2016).
16. Dunkin, C. S. et al. Scarring occurs at a critical depth of skin injury: precise measurement in a graduated dermal scratch in human volunteers. *Plast. Reconstr. Surg.* **119**, 1722–1734 (2007).
17. Koehler, R. H. et al. Minimal adhesions to ePTFE mesh after laparoscopic ventral incisional hernia repair: reoperative findings in 65 cases. *JSLs* **7**, 335–340 (2003).
18. Rippa, A. L., Kalabusheva, E. P., & Vorotelyak, E., A. Regeneration of dermis: scarring and cells involved. *Cells* **8**, 607 (2019).
19. Zamir, E. A., Rongish, B. J. & Little, C. D. The ECM moves during primitive streak formation—computation of ECM versus cellular motion. *PLoS Biol.* **6**, e247 (2008).
20. Szabó, A., Rupp, P. A., Rongish, B. J., Little, C. D. & Czirók, A. Extracellular matrix fluctuations during early embryogenesis. *Phys. Biol.* **8**, 045006 (2011).
21. Aleksandrova, A. et al. Convective tissue movements play a major role in avian endocardial morphogenesis. *Dev. Biol.* **363**, 348–361 (2012).
22. Loganathan, R. et al. Extracellular matrix motion and early morphogenesis. *Development* **143**, 2056–2065 (2016).
23. Miron-Mendoza, M., Koppaka, V., Zhou, C. & Petroll, W. M. Techniques for assessing 3-D cell-matrix mechanical interactions in vitro and in vivo. *Exp. Cell Res.* **319**, 2470–2480 (2013).
24. Sakar, M. S. et al. Cellular forces and matrix assembly coordinate fibrous tissue repair. *Nat. Commun.* **7**, 11036 (2016).
25. Abu-Hijleh, M. F., Roshier, A. L., Al-Shboul, Q., Dharap, A. S. & Harris, P. F. The membranous layer of superficial fascia: evidence for its widespread distribution in the body. *Surg. Radiol. Anat.* **28**, 606–619 (2006).
26. Avelar, J. Regional distribution and behavior of the subcutaneous tissue concerning selection and indication for liposuction. *Aesthetic Plast. Surg.* **13**, 155–165 (1989).
27. Lockwood, T. E. Superficial fascial system (SFS) of the trunk and extremities: a new concept. *Plast. Reconstr. Surg.* **87**, 1009–1018 (1991).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Mice and genotyping

All mouse strains (C57BL/6J, *En1^{cre}*, *R26^{VT2/GK3}*, *R26^{mTmG}*, *R26^{idTR}*, *Rag2^{-/-}* and Fox Chase SCID) were obtained from either Jackson laboratories, Charles River or generated at the Stanford University Research Animal Facility as described previously¹². Animals were housed at the Helmholtz Center Animal Facility. Cages were maintained at constant temperature and humidity with a 12-h light cycle. Animals were supplied with food and water ad libitum. All animal experiments were reviewed and approved by the Government of Upper Bavaria and registered under the projects 55.2-1-54-2532-16-61 and 55.2-2532-02-19-23 and conducted under strict governmental and international guidelines. This study is compliant with all relevant ethical regulations regarding animal research. Unless specified, 8-to-10-week-old adult mice were used for the animal experiments. Both male and female mice were used. Cre-positive (*Cre⁺*) animals from double-transgenic reporter mice were identified by detection of relevant fluorescence in the dorsal dermis. Genotyping was performed to distinguish mouse lines containing a 200-base pair (bp) *cre* fragment (*cre^{+/+}*) from the wild-type (*cre^{-/-}*). Genomic DNA from the ear-clips was extracted using QuickExtract DNA extraction solution (Epicentre) following the manufacturer's guidelines. DNA extract (1 µl) was added to each 24 µl PCR. The reaction mixture was set up using Taq PCR core kit (Qiagen) containing 1× coral buffer, 10mM dNTPs, 0.625 units Taq polymerase, 0.5 µM forward primer 'Cre_genotype_4F' 5'-ATTGCTGTCACCTGGTCGTGGC-3' (Sigma) and 0.5 µM reverse primer 'Cre_genotype_4R' 5'-GGAAAATGCTTCTGTCCGTTTGC-3' (Sigma). PCRs were performed with initial denaturation for 10 min at 94 °C, amplification for 30 cycles (denaturation for 30 s at 94 °C, hybridization for 30 s at 56 °C, and elongation for 30 s at 72 °C) and final elongation for 8 min at 72 °C, and then cooled to 4 °C. In every experiment, negative controls (non-template and extraction) and positive controls were included. The reactions were carried out in an Eppendorf master cycler. Reactions were analysed by gel electrophoresis.

Human skin samples

Fresh human skin and scar biopsies, from various anatomic locations, were collected from donors between 18–65 years of age, through the Section of Plastic and Aesthetic Surgery, Red Cross Hospital Munich (reference number 2018-157), and by the Department of Dermatology and Allergology, Klinikum rechts der Isar Technical University Munich (reference number 85/18S). Informed consent was obtained from all subjects before skin biopsies. Upon collection, these samples were directly processed for tissue culture or fixed with PFA and then processed for cryosection or paraffin section followed by histological or immunofluorescent analyses.

Fascia in vitro culture

Two in vitro systems were used. To visualize the changes in matrix architecture in real time, 2-mm-diameter biopsies were excised from PO C57BL/6J neonates and processed for live imaging (SCAD assay, Patent Application no. PLA17A13). To determine the effectiveness of the DT treatment, muscle and fascia were manually separated from the rest of the skin in the chimeric grafts experiments and incubated with DT at different concentrations for 1 h at ambient temperature. Next, samples were washed with PBS and incubated in DMEM/F12 (Thermo Fisher) supplemented with 10% serum (Thermo Fisher), 1% penicillin–streptavidin (Thermo Fisher), 1% GlutaMAX (Thermo Fisher) and 1% non-essential amino acids solution (Thermo Fisher) in a 37 °C, 5% CO₂ incubator. Medium was routinely exchanged every other day. Samples were fixed at day 6 of culture with 2% paraformaldehyde and processed for histology.

Histology

Tissue samples were fixed overnight with 2% paraformaldehyde in PBS at 4 °C. Samples were rinsed three times with PBS, embedded in optimal

cutting temperature (OCT, Sakura Finetek) and flash-frozen on dry ice. Six-micrometre sections were made in a Cryostar NX70 cryostat (Thermo fisher). Masson's trichrome staining was performed with a Sigma-Aldrich trichrome stain kit, according to the manufacturer's guidelines. For immunolabelling, sections were air-dried for 5 min and fixed with –20 °C-chilled acetone for 20 min. Sections were rinsed three times with PBS and blocked for 1 h at room temperature with 10% serum in PBS. Then, the sections were incubated with primary antibody in blocking solution for 3 h at ambient temperature. Sections were then rinsed three times with PBS and incubated with secondary antibody in blocking solution for 60 min at ambient temperature. Finally, sections were rinsed three times in PBS and mounted with fluorescent mounting media with 4,6-diamidino-2-phenylindole (DAPI). Primary antibodies used: goat-anti-ACTA2(αSMA) (1:50, Abcam), rabbit-anti-TUBB3 (1:100, Abcam), rat-anti-THY1(CD90) (1:100, Abcam), rat-anti-CD24 (1:50, BD biosciences), rabbit-anti-DPP4(CD26) (1:150, Abcam), rabbit-anti-PECAM1 (1:10, Abcam), rat-anti-CD34 (1:100, Abcam), rabbit-anti-collagen I (1:150, Rockland), rabbit-anti-collagen III (1:150, Abcam), rabbit-anti-collagen VI (1:150, Abcam), rabbit-anti-DLK1 (1:200, Abcam), rat-anti-ERT7 (1:200, Abcam), rat-anti-F4/80 (1:400, Abcam), rabbit-anti-LYVE1 (1:100, Abcam), rat-anti-MOMA2 (1:100, Abcam), goat-anti-PDGFRα (1:50, R & D systems), rat-anti-LY6A(Sca1) (1:150, Biolegend), rat-anti-CD44 (1:100, Abcam), rabbit-anti-NOV/CN3 (1:20, Elabscience), sheep-anti-FAP (1:100, R&D systems), rat-anti-IL12 (1:50, Biolegend), rat-anti-IL4 (1:50, Biolegend), rat-anti-CD19 (1:20, BD biosciences), hamster-anti-CD3e (1:100, Biolegend), rat-anti-NCR1 (1:20, Biolegend) and rat-anti-LY6G (1:100, Abcam). PacificBlue-, Alexa Fluor 488-, Alexa Fluor 568- or Alexa Fluor 647-conjugated antibodies (1:500, Life technologies) against suitable species were used as secondary antibodies.

Microscopy

Histological sections were imaged using a ZEISS AxioImager. Z2m (Carl Zeiss). For whole-mount 3D imaging of wounds, fixed samples were embedded in 35-mm glass bottom dishes (Ibidi) with low-melting point agarose (Biozym) and left to solidify for 30 min. Imaging was performed using a Leica SP8 multi photon microscope (Leica, Germany). For live imaging of fascia cultures, samples were embedded as just above. Attention was paid to mount the samples with the fascia facing up towards the objective. Imaging medium (DMEM/F-12; SiR-DNA 1:1,000) was then added. Time-lapse imaging was performed over 20 h under the multi-photon microscope. A modified incubation system, with heating and gas control (ibidi, catalogue nos. 10915 and 11922), was used to guarantee physiologic and stable conditions during imaging. Temperature control was set to 35 °C with 5% CO₂-supplemented air. Second harmonic generation signal and green auto-fluorescence as a reference were recorded every hour. 3D and 4D data was processed with Imaris 9.1.0 (Bitplane) and ImageJ (1.52i). Contrast and brightness were adjusted for better visibility.

Image analysis

Histological images were analysed using ImageJ. For quantification of labelled cells in our fate mapping experiments, we manually defined the wound, surrounding dermis, and adjacent fascia areas. We defined the wound as the area flanked by the near hair follicles on both sides, extending from the base of the epidermis down to the level of the hair follicles bulges. Surrounding dermis area was defined as the 200 µm immediately adjacent to the wound on both sides. Fascia area was defined as the tissue immediately below the wound. The number of labelled cells in each area was determined by quantifying the particles that were double-positive for DAPI and for the desired label (for example, DiI or GFP) channels. The coverage of the labelled matrix in the wound area was determined by quantifying the area that was double-positive for the labelled matrix and the collagen I-, III- and VI-staining signal. Cell density of *En1^{cre}*, *R26^{idTR}* cultures treated with DT was

Article

quantified by dividing the total cells (DAPI) by the matrix area (collagen I, III and VI). Collagens density was calculated as the collagens area coverage of the entire section area. Matrix movements in live-imaged cultures were determined by tracking the length of the two furthest points along the sample in both the SHG and auto-fluorescence channels. Length measurements were normalized to the original length at time 0. Wound size was normalized for each time point using the original area at day 0. Scar length was quantified from randomly selected sections taken from the middle of the scar using the two flanking hair follicles as a reference. RFI was calculated by measuring the mean grey value and normalize to the dermis images. Fractal analysis was performed using the ImageJ plug-in 'FracLac'29 (FracLac2015Sep090313a9390) using the same settings and preprocessing as previously described¹³.

Dil labelling of fascia in animals

Two 5-mm-diameter full-thickness excisional wounds were created on the back of 8- to 10-week-old C57BL6/J mice with a biopsy punch. Lipophilic Vybrant Dil dye (Life technologies, V22885) (10–20 μ l) was injected into the exposed fascia directly above the dorsal muscles. Wounded tissue was harvested 14 dpw and processed for histology and imaging by fluorescence microscopy.

Chimeric skin transplantation

Full-thickness 6-mm-diameter biopsies were collected from the back-skin of either $R26^{mTmG}$, $R26^{VT2/GK3}$, $En1^{cre};R26^{mTmG}$, $En1^{cre};R26^{IDTR}$ or C57BL6/J adult mice. Using the PC muscle layer as an anatomical reference, the fascia together with the muscle layer were carefully separated from the dermis and epidermis using Dumont no. 5 forceps (Fine Science Tools) and a 26G needle under the fluorescent stereomicroscope (M205 FA, Leica). EPFs from fascia + muscle samples of $En1^{cre};R26^{IDTR}$ mice were ablated by incubation with 20 μ g ml⁻¹ of diphtheria toxin (Sigma-Aldrich, D0564) or only DMEM/F12 as vehicle for 1 h at ambient temperature followed by 3 washing steps with PBS. At this point, the matrix samples were labelled by incubation with 100 μ M Alexa Fluor NHS Ester (Life Technologies, A20006) or Pacific Blue succinimidyl ester (Thermo Fisher, P10163) in PBS for 1 h at ambient temperature followed by 3 washing steps with PBS. Chimeras were made by placing the epidermis + dermis portion of a mouse strain on top of the muscle + fascia of another strain and left to rest for 20 min at 4 °C inside a 35-mm culture dish with 2 ml of DMEM/F12. Special attention was paid on preserving the original order of the different layers (top to bottom: epidermis > dermis > muscle > fascia). Then, a 2 mm 'deep' full thickness was excised from the chimeric graft using a biopsy punch in the middle of the biopsy. To create 'superficial' wounds, the 2-mm excision was done only in the epidermis + dermis half before reconstitution with the bottom part. 'Wounded' chimeric grafts were then transplanted into freshly-made 4-mm-diameter full-thickness excisional wounds in the back of either $RAG2^{-/-}$ or Fox Chase SCID immunodeficient 8- to 10-week-old mice. Precautions were taken to clean out the host blood from the fresh wound before the transplant and to leave the graft to dry for at least 20 min before ending the anaesthesia, to increase the transplantation success. To prevent mice from removing the graft, a transparent dressing (Tegaderm, 3M) was placed on top of the grafts.

In situ matrix tracing and EdU pulses

Eight- to ten-week-old C57BL6/J mice received subcutaneous 20- μ l injections of 10 mg ml⁻¹ FITC NHS ester in physiological saline with 0.1 M sodium bicarbonate pH9 (46409, Life technologies) four and two days before wounding. At 2, 6 or 13 dpw, mice received 200 μ l intraperitoneal injections of 1 mg ml⁻¹ EdU in PBS. Samples were collected 24 h after the EdU pulse and processed for cryosection and imaging by fluorescence microscopy.

Flow cytometry

Fascia and dermis were physically separated from the back-skin of C57BL6/J or $En1^{cre};R26^{mTmG}$ mice under the fluorescence

stereomicroscope as before. Harvested tissue was minced with surgical scissors and digested with an enzymatic cocktail containing 1 mg ml⁻¹ collagenase IV, 0.5 mg ml⁻¹ hyaluronidase, and 25 U ml⁻¹ DNase I (Sigma-Aldrich) at 37 °C for 30 min. The resulted single cell suspension was filtered and incubated with conjugated/unconjugated primary antibodies (dilution 1:200) at 4 °C for 30 min, followed by an incubation with a suitable secondary antibody when needed at 4 °C for 30 min. Cells were washed and stained with Sytox blue dye (dilution 1:1,000, Life technologies, S34857) for dead cell exclusion. Cells were subjected to flow cytometric analysis using a FACS Aria III (BD Bioscience). Fibroblasts were gated as the lineage negative (Lin⁻) fraction not expressing the PTPRC, PECAM1, LY76, LYVE-1, or EPCAM markers. Primary antibodies used: anti-DLK1 (Abcam), anti-CD9 (Santa Cruz), anti-NGFR (Miltenyi), anti-F4/80 (Abcam), AlexaFluor790-anti-NG2 (Santa Cruz), FITC-anti-DPP4 (eBioscience), PerCP-eFluor710-anti-ITGB1 (eBioscience), anti-CD34 (Abcam), PerCP-Cy5.5-anti-CD24 (eBioscience), APC-Fire750-anti-CD34 (Biolegend), APC-anti-ITGA7 (R&D systems), PerCP-Cy5.5-anti-LY6A (eBioscience), PE-Vio770-anti-PDGFR α (Miltenyi), PerCP-Vio700-anti-CD146 (Miltenyi), APC-anti-PECAM1 (eBioscience), eFluor660-anti-LYVE1 (Thermo fisher), APC-LY76 (TER119), APC-anti-EPCAM (CD326) and APC-anti-PTPRC. Secondary antibodies used: Alexa Fluor 488 goat anti-rabbit (Life technologies) and AlexaFluor568 goat anti-rat (Life technologies).

Scanning electron microscopy

Skin biopsies of adult C57BL6/J mice were collected, and the fascia was manually separated as before. Samples were then fixed overnight with paraformaldehyde and glutaraldehyde, 3% each, in 0.1% sodium cacodylate buffer pH 7.4 (Electron Microscopy Sciences). Samples were dehydrated in gradual ethanol and dried by the critical-point method, using CO₂ as the transitional fluid (Polaron Critical Point Dryer CPC E3000; Quorum Technologies) and observed by scanning electron microscopy (JSM 6300F; JEOL).

ePTFE membrane implants

These membranes are routinely used in the clinic to circumvent post-operative adhesions after laparoscopic ventral incisional hernia repairs¹⁷. Two 6-mm-diameter full-thickness excisional wounds were created with a biopsy punch on the back of 8-week-old $En1^{cre};R26^{VT2/GK3}$ or C57BL6/J mice. Sterile 8-mm-diameter ePTFE impermeable membranes (Dualmesh, GORE) were implanted between the surrounding skin and the fascia underneath, to cover the open wound on the right side. For this, the surrounding skin was loosen using Dumont no. 5 forceps and spatula (10090-13, Fine Science Tools). The dual-surface membrane was implanted with the attaching face facing out, so to promote dermal cell attachment, while the smooth surface was in direct contact with the fascia. The left sham control wound underwent the same procedure without implanting any membrane. Each wound was photographed at indicated time points, and wound areas were measured using ImageJ. Wound sizes at any given time point after wounding were expressed as percentage of initial (day 0) wound area. At 7 or 63 dpw, samples were collected and processed for histology.

Released fascia injury in adult mice

Two 5-mm-diameter full-thickness excisional wounds were created with a biopsy punch on the back of 8-week old male C57BL6/J mice. The skin around the wound on the left side was separated from the underneath skeletal muscle using a sterilized gold-plated 3 \times 5 mm genepaddles (Harvard Apparatus, 45-0122) to release the fascia layer. The right wound served as a control. Each wound was digitally photographed at indicated time points, and wound areas were measured using Photoshop (Adobe Systems). Wound sizes at any given time point after wounding were expressed as percentage of initial (day 0) wound area. The harvested tissue at the indicated time points was processed for cryosection and Masson's trichrome staining for histology.

Viral particle production

AAV6 expressing GFP or Cre recombinase were produced by transfecting the AAVpro 293T Cell Line (Takara Bio, 632273) with pAAV-U6-sgRNA-CMV-GFP (Addgene, 8545142) or pAAV-CRE Recombinase vector (Takara Bio, 6654), pRC6 and pHelper plasmids procured from AAVpro Helper Free System (Takara Bio, 6651). Transfection was performed with PEI transfection reagent and viruses were harvested 72 h later and purified with an AAVpro purification kit (Takara Bio, 6666) and titre was calculated using real-time PCR.

Fascia cells ablation with AAV6-Cre viral particles and DT treatment in pups

Two 3-mm-diameter full-thickness excisional wounds were created with a biopsy punch on the back of postnatal day 11 (P11) *R26^{idTR}* mice. Twenty microlitres of Cre-expressing AAV6-Cre or control AAV6-eGFP at viral titre of $5 \times 10^{11} \text{ ml}^{-1}$ were injected subcutaneously at the area between the two wounds. DT solution at $1 \text{ ng } \mu\text{l}^{-1}$ in PBS was intraperitoneally injected to each mouse once per day for 7 days at the dosage of 5 ng g^{-1} . Tissue was collected seven days after wounding.

Statistics

Statistical analyses were performed using GraphPad Prism software (v.6.0, GraphPad). Statistical test and *P* values are specified in the figure legends and in the corresponding plots. For simplicity, *P* values below 0.0001 were stated as equal to 0.0001.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source Data for Figs. 1–5 and Extended Data Figs. 1–10 are provided with the paper. Additional information is available from the corresponding author on reasonable request.

Acknowledgements We thank G. Mettenleiter for technical assistance with the SEM; S. Dietzel and the Core Facility Bioimaging at the Biomedical Centre of the Ludwig-Maximilians-Universität München for access and support with the multi-photon system. Y.R. was supported by the Human Frontier Science Program Career Development Award (CDA00017/2016), the German Research Foundation (RI 2787/1-1 AOBJ: 628819), the Fritz-Thyssen-Stiftung (2016-01277) and the European Research Council Consolidator Grant (ERC-CoG 819933). D.C.-G. was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Deutscher Akademischer Austauschdienst (DAAD).

Author contributions Y.R. outlined and supervised the research narrative and experimental design. D.C.-G. performed the cell and matrix tracing experiments, chimeric skin transplantations, ePTEF implantations and image analysis. D.J. performed the cell depletion with DT in pups, flow cytometry experiments and analysis, fascia release experiments and experiments with human skin samples. S.C. performed the live imaging and analysis. S.C. and P.R. performed 3D imaging and analysis. H.Y. performed histology and immunofluorescence staining of human samples. J.W. provided veterinary advice and prepared animal experiment protocols. S.K.G. made the viral particles and performed titre quantification. Q.Y. assisted in the flow cytometry experiments. U.M. and T.V. collected consent from patients and the primary human tissue samples, and assisted with translational and clinical advice. U.M. wrote, in part, the ethical application for collection and use of human samples. M.A. and A.W. performed the scanning electron microscopy images. Y.R., D.C.-G. and D.J. wrote the manuscript.

Competing interests The authors declare no competing interests.

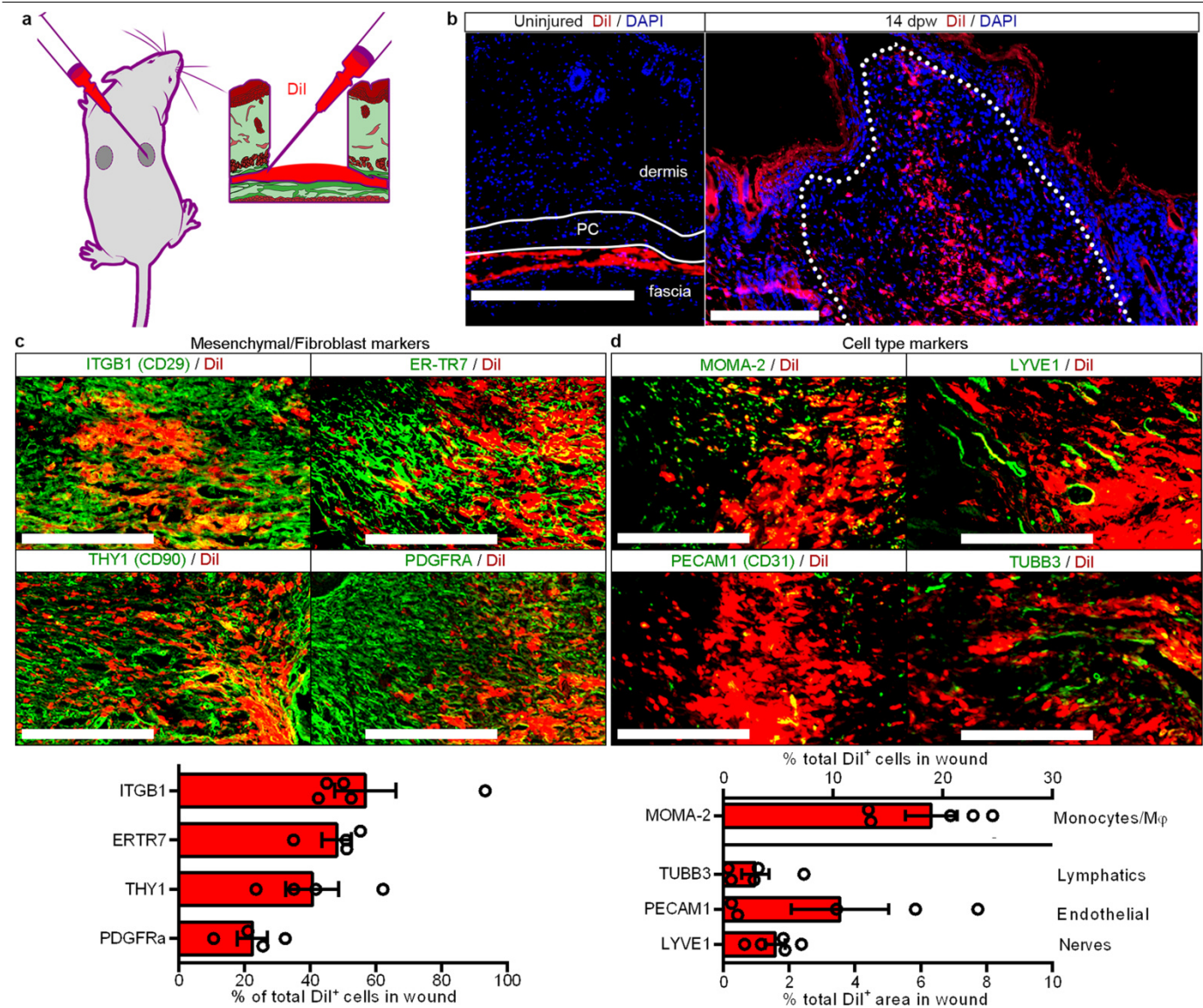
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1794-y>.

Correspondence and requests for materials should be addressed to Y.R.

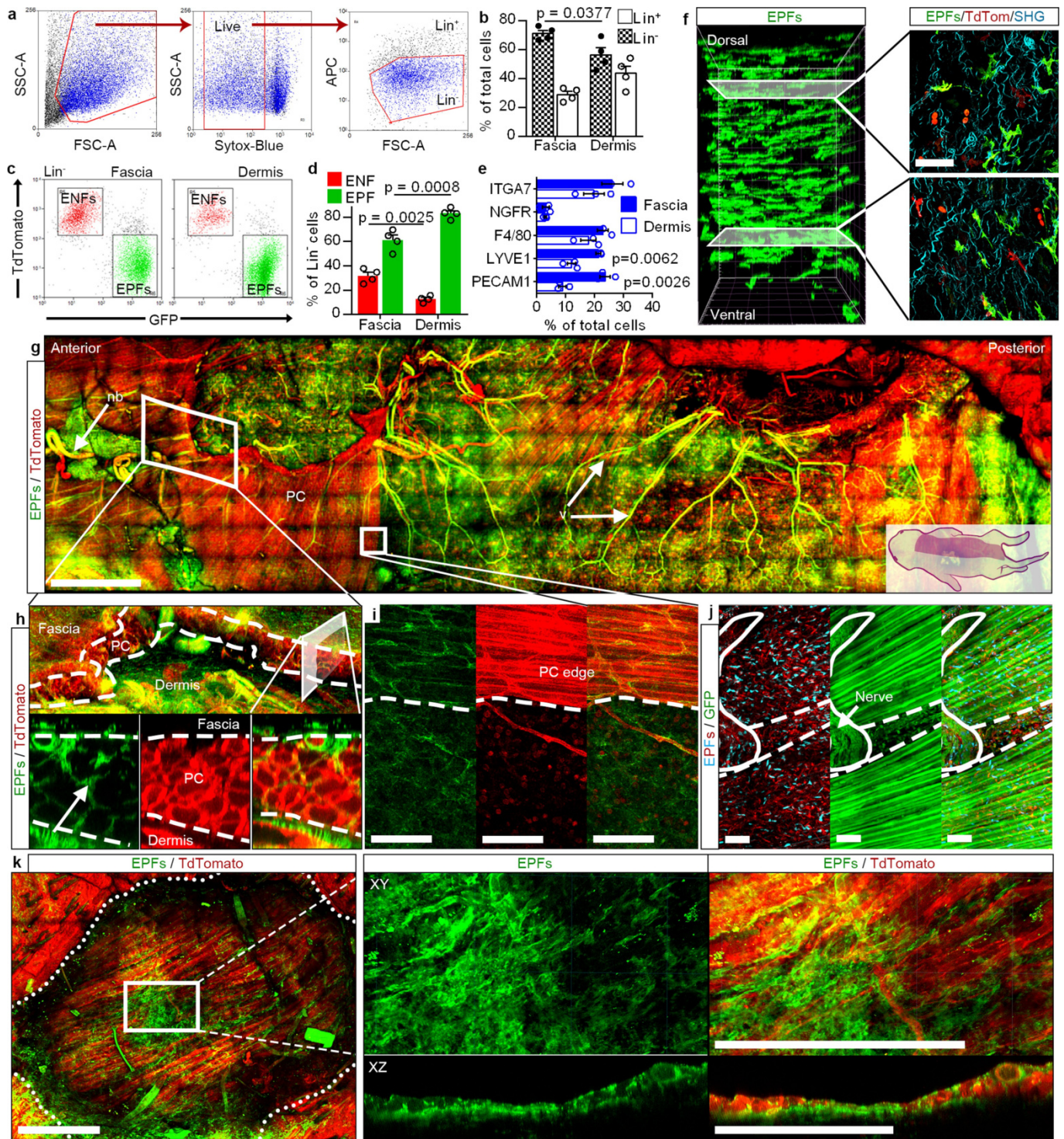
Peer review information Nature thanks Christopher Buckley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



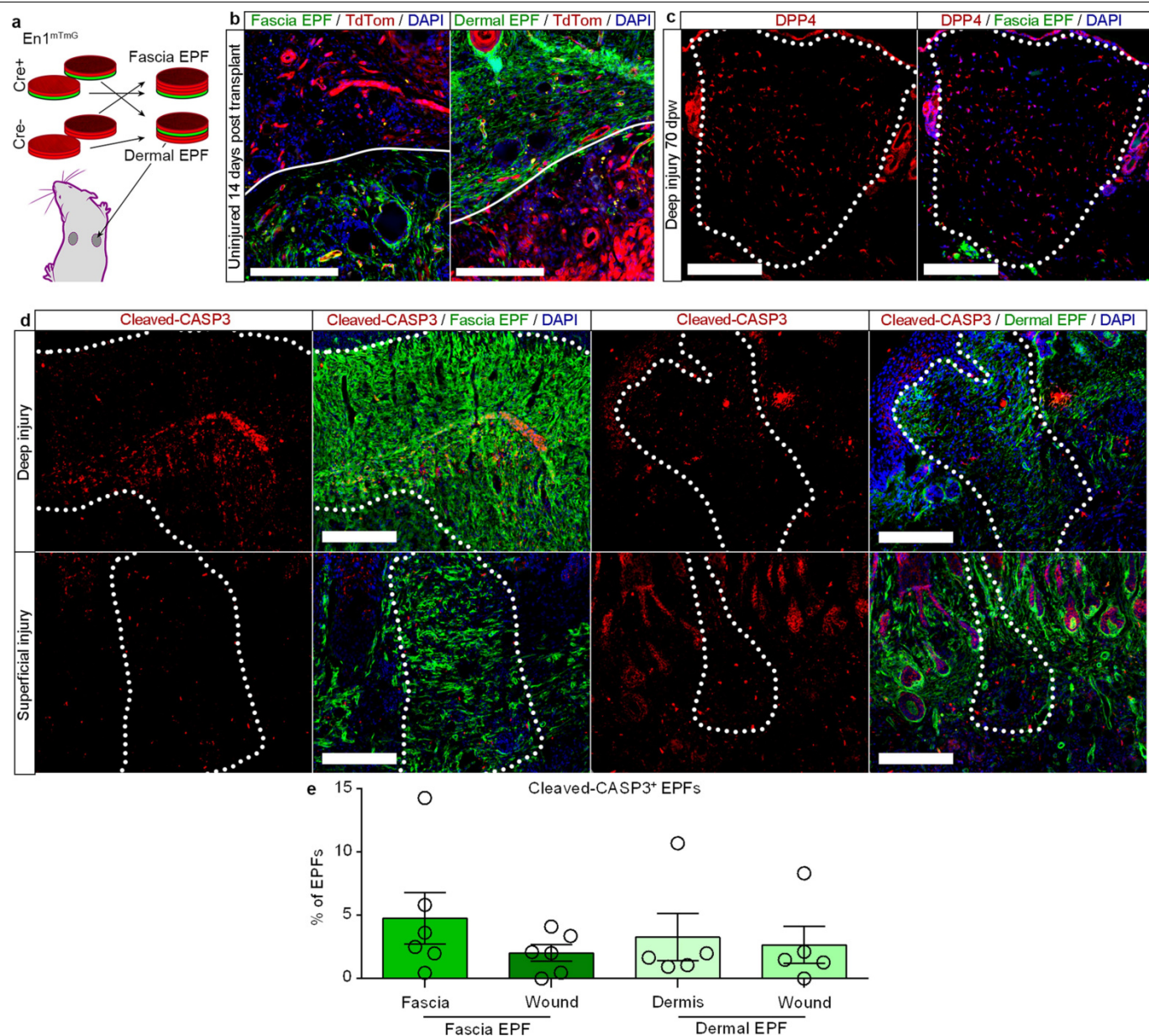
Extended Data Fig. 1 | Fate mapping of fascia cells with Dil. **a**, Dil labelling of fascia cells. **b**, Histology showing Dil⁺ cells in uninjured controls (left) and at 14 dpw (right). Representative images of five biological replicates. **c**, Immunolabelling (top) and fractions (bottom) of Dil-positive cells expressing mesenchymal/fibroblast markers ITGB1, ER-TR7, THY1 and PDGFRα. Data are mean ± s.e.m.; *n* = 4 (5 in ITGB1) images analysed from 3

biological replicates. **d**, Immunolabelling (top) and fractions (bottom) of Dil-positive monocytes/macrophages (MOMA-2), lymphatic cells (LYVE1), endothelial cells (PECAM1) and nerve cells (TUBB3). Data are mean ± s.e.m.; *n* = 5 images analysed from 3 biological replicates. Lines delimit PC. The dotted line delimits the wound. Scale bars, 200 μm.



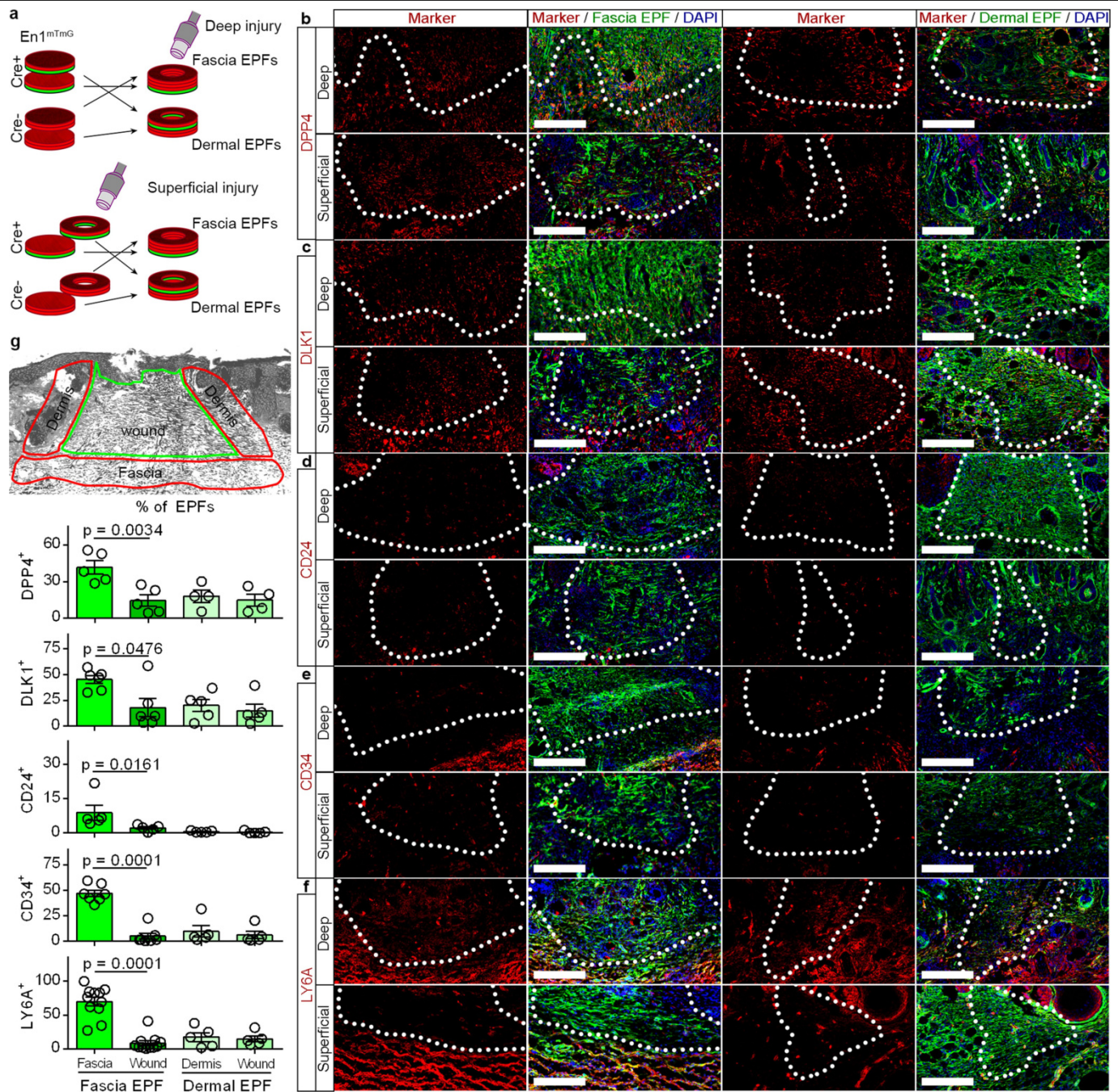
Extended Data Fig. 2 | Fascia EPFs traverse PC. **a**, Gating strategy for fibroblasts analysis. **b**, Percentages of fibroblasts (Lin⁻) and lineage-positive cells in fascia and dermis. Data are mean \pm s.e.m.; $n = 4$ independent experiments. Unpaired two-tailed t -test, 95% CI. **c**, Scatter plots of EPFs (GFP⁺Lin⁻) and ENFs (TdTomato⁺Lin⁻) in fascia and dermis. Representative plots of three independent experiments. **d**, EPF and ENF fractions in fascia and dermis. Data are mean \pm s.e.m.; $n = 4$ independent experiments. Two-way ANOVA, multiple comparison Tukey's test, 95% CI. **e**, Endothelial cell (PECAM1⁺), lymphatic cell (LYVE1⁺), macrophages (F4/80⁺) and nerve cell (NGFR⁺) fractions in fascia and dermis. Data are mean \pm s.e.m.; $n = 3$ biological replicates. Two-way ANOVA, multiple comparison Tukey's test, 95% CI. **f-k**, Representative images

of 3D-rendered *En1^{cre};R26^{mTmG}* or *En1^{cre};R26^{VT2/Gk3}* back skin fascia from at least three biological replicates. **f**, Lateral view (left) and cross-sections (right) of adult fascia. **g**, Top view (ventral side up) of neonate back skin. **h**, Top side view (top) and lateral cross-section (bottom) at the forelimb junction showing EPF traversing the PC. **i**, Top view at a muscle breach showing EPFs in both locations. **j**, Top view at a muscle opening where nerves pass through and polyclones of EPFs reside. **k**, Top view (top, epidermis side up) and lateral cross-section (bottom) of an adult superficial wound (3 dpw). Broken lines delimit PC. Dotted lines delimit the epidermis. Scale bars, 1,500 μ m (**g**), 100 μ m (**f**, **i**, **j**) and 500 μ m (**k**); v, vessels; nb, nerve bundles.



Extended Data Fig. 3 | Fascia EPFs maintain position in steady conditions and recede from wounds over time. **a**, Dermal versus fascia EPFs chimeras in uninjured conditions. **b**, Fascia (left) or dermal (right) EPFs-traced chimeras. Representative images of 3 biological replicates. **c**, Scars at 70 dpw from deep injuries of fascia EPF-traced chimeras immunolabelled for DPP4. Representative images of 3 biological replicates. **d**, Cleaved CASP3 expression

in wounds from fascia (left) or dermal (right) EPF-traced chimeras from deep (top) or superficial (bottom) injuries at 14 dpw. **e**, Fractions of fascia or dermal EPFs in the wound, dermis or fascia control regions positive for cleaved CASP3. Data are mean \pm s.e.m.; $n = 6$ and 5 (fascia and dermal EPF, respectively) images analysed from 5 biological replicates. Lines delimit the border between fascia and dermis. Dotted lines delimit the wound or scar. Scale bars, 200 μ m.



Extended Data Fig. 4 | Fascia EPFs express wound fibroblast markers.

a, Dermal versus fascia EPF-traced chimeras with two injury conditions.

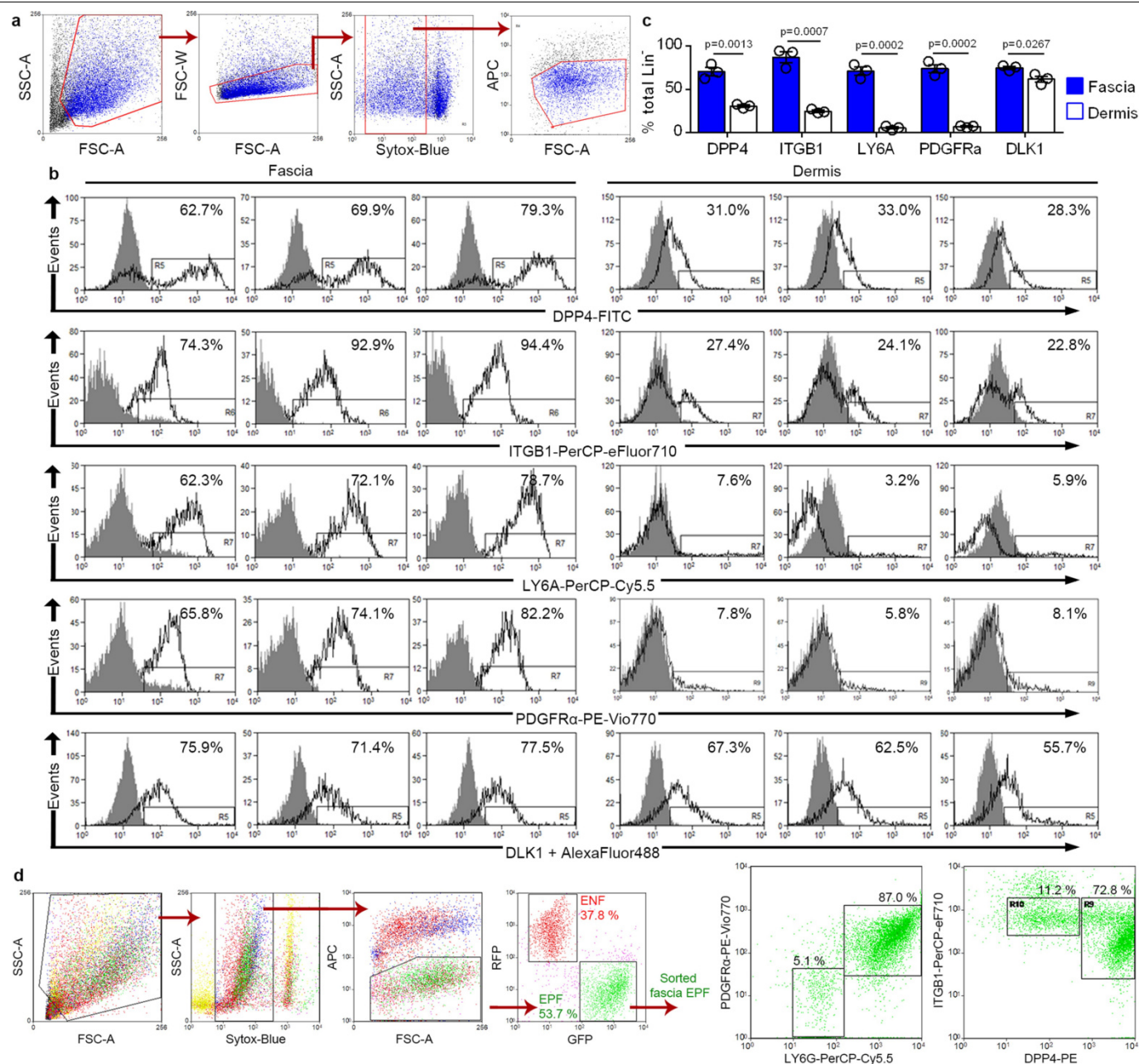
b–f, Representative immunolabelling for the fibroblast markers DPP4 (**b**), DLK1

(**c**), CD24 (**d**), CD34 (**e**) and LY6A (**f**) from 4 biological replicates. **g**, Areas

analysed (top) for marker-positive EPF quantification (bottom). Data are

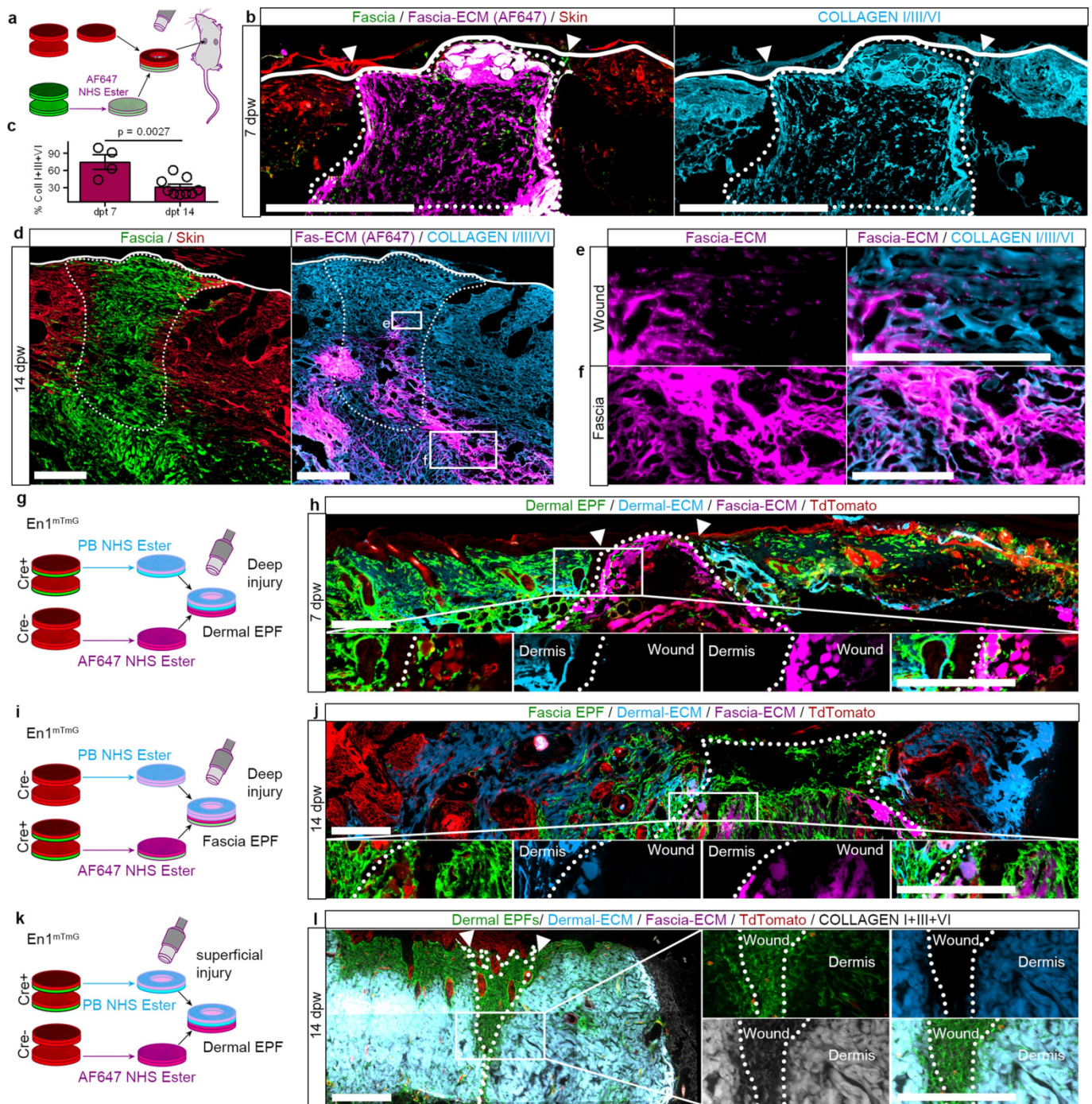
mean \pm s.e.m.; $n = 4$ (DPP4 in dermal EPFs), 5 (all markers in dermal EPFs with the

exception of DPP4 and all markers in fascia EPFs with the exception of DLK1, CD34 and LY6A), 6 (DLK1 in fascia EPFs), 7 (CD34 in fascia EPFs) or 11 (LY6A in fascia EPFs) images analysed from 4 biological replicates. One-way ANOVA, multiple comparison Tukey's test, 95% CI. Dotted lines delimit the wound bed. Scale bars, 200 μ m.



Extended Data Fig. 5 | Differential expression of classical markers on fascia and dermal fibroblasts. a, Gating strategy for fibroblast (Lin⁻) cytometry. **b,** Histogram plots of fibroblast-marker expression in fascia or dermis derived fibroblasts from three biological replicates. **c,** Fraction of marker-positive cells

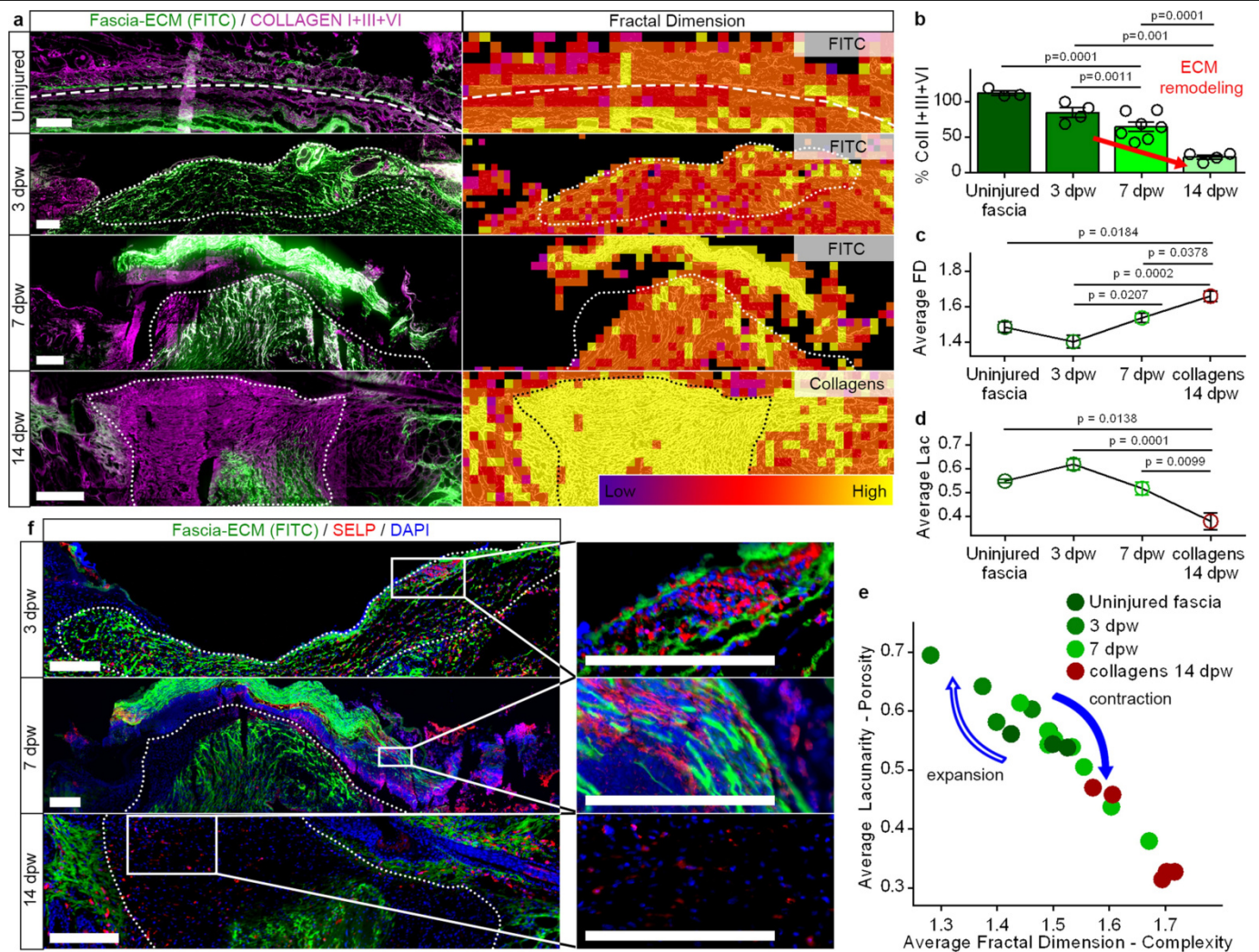
from total fibroblast population. Data are mean \pm s.e.m.; $n = 3$ biological replicates. Unpaired two-tailed t -test, 95% CI. **d,** Gating strategy for fascia EPF (Lin⁻ GFP⁻) sorting and detection of LY6A, PDGFR1, DPP4 and ITGB1 expression. Representative plots of three biological replicates.



Extended Data Fig. 6 | Fascia but not dermal matrix steers into wounds.

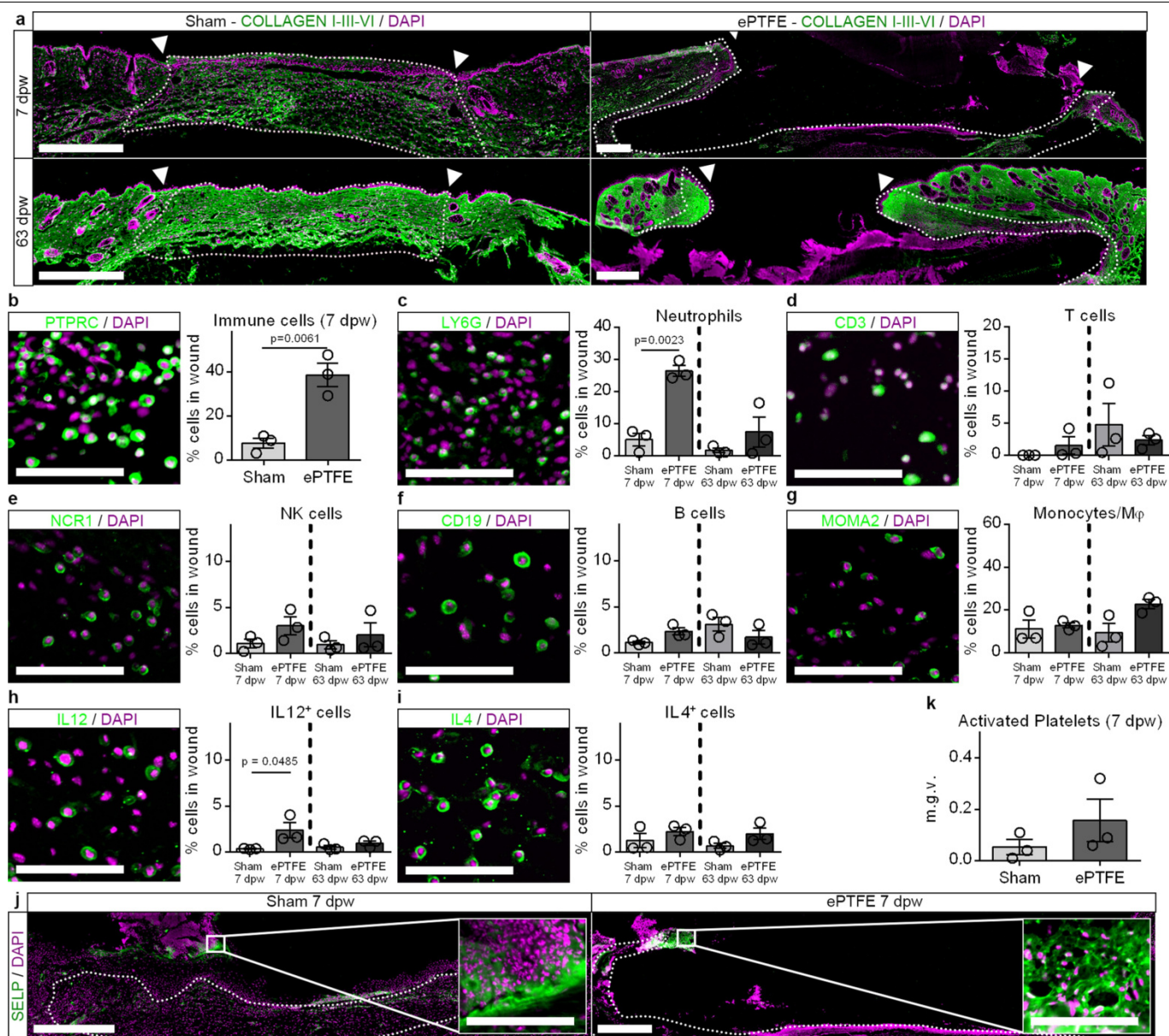
a, Matrix tracing in chimeric grafts. **b**, Grafts at 7 dpw immunolabelled for collagen I, III and VI. Representative image of three biological replicates. **c**, Label coverage fraction from total collagens in the wound at defined time points. Data are mean \pm s.e.m.; $n = 4$ (7 dpw) and 9 (14 dpw) sections analysed from 3 biological replicates. Unpaired two-tailed t -test, 95% CI. **d**, Wounds at 14 dpw immunolabelled for collagens. Representative images of three biological replicates. **e-f**, Higher magnification of the insets in **d**. **g**, Experimental design for deep injury. **h**, Double matrix tracing in deep-injured dermal EPF-traced grafts at 7 dpw.

Representative image of three biological replicates. **i,j**, Double matrix tracing in deep-injured fascia EPF-traced grafts at 14 dpw. Representative image of three biological replicates. **k,l**, Double matrix labelling in superficial-injured dermal EPF-traced grafts at 14 dpw immunolabelled for collagens. Representative image of three biological replicates. Dotted lines delimit the wound. Arrowheads mark the original injury. Continuous lines delimit the epidermis dermis margin. Scale bars, 500 μ m (**b**), 100 μ m (**d-f**) and 200 μ m (**h,j,l**).



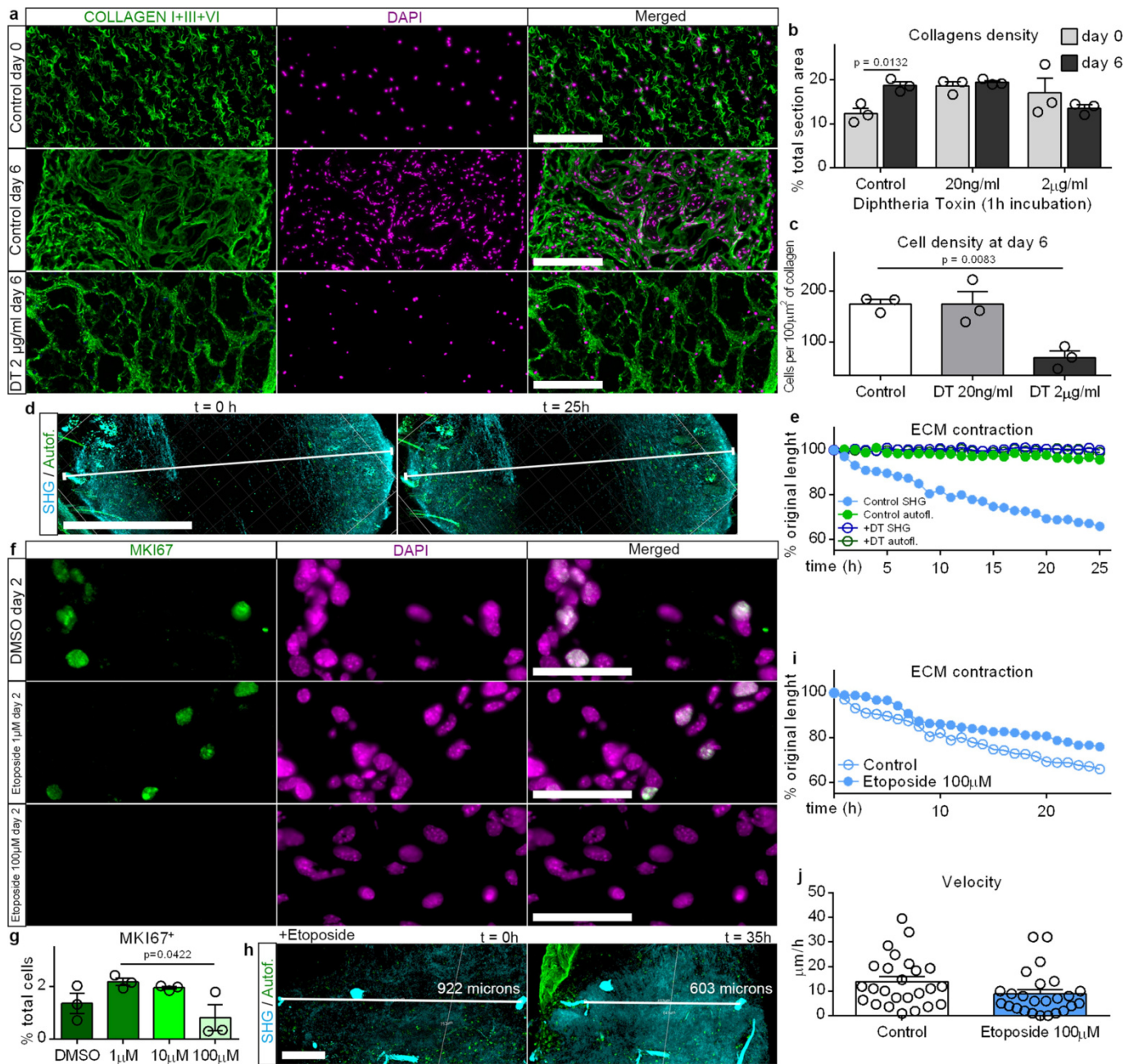
Extended Data Fig. 7 | Fascia matrix forms the eschar and is remodelled in the wound. a, Left, in situ matrix tracing and collagen I, III and VI immunolabelling at defined time points after wounding. Representative images of three biological replicates. Right, subsampled fractal dimension maps of the FITC signal of uninjured tissue and at 3 and 7 dpw, and the collagen signal at 14 dpw. **b**, Matrix label coverage from total collagen I, III and VI signal in the wound. Data are mean \pm s.e.m.; $n = 3$ (uninjured), 4 (3 dpw), 7 (7 dpw) and 4 (14 dpw) sections analysed from 3 biological replicates. One-way ANOVA,

Tukey's multiple comparisons. **c**, **d**, Average fractal dimension (**c**) and lacunarity (**d**) from subsampled maps. Data are mean \pm s.e.m.; $n = 5$ (uninjured), 5 (3 dpw), 8 (7 dpw) and 3 (14 dpw) images analysed from 3 biological replicates. One-way ANOVA, Tukey's test, 95% CI. **e**, Scatter plot of average fractal dimension and lacunarity values. **f**, In situ matrix tracing and SELP immunolabelling at defined time points after wounding. Representative images of three biological replicates. The broken line separates dermis from fascia. Dotted lines indicate the wound. Scale bars, 200 μ m.



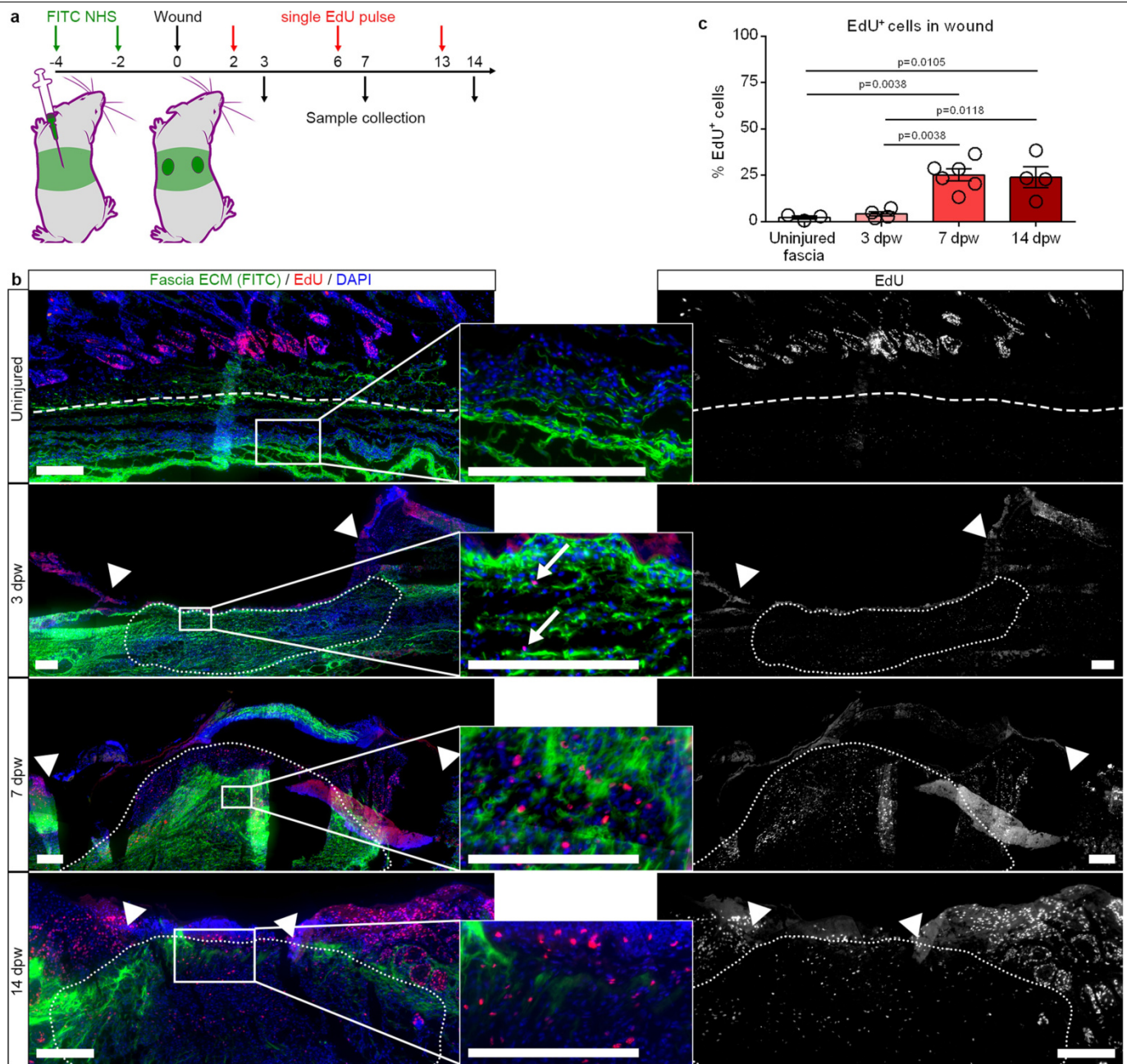
Extended Data Fig. 8 | Inflammation resolution and coagulation stay unaffected during fascia blocking. **a**, Sham (left) or ePTFE-implanted (right) wounds at 7 (top) or 63 dpw (bottom) immunolabelled for collagen I, III and VI. Representative images of three biological replicates. **b–i**, Immunolabellings (left) and fractions (right) of immune cells (PTPRC $^{+}$; **b**), neutrophils (LY6G $^{+}$; **c**), T cells (CD3 $^{+}$; **d**), NK cells (NCR1 $^{+}$; **e**), B cells (CD19 $^{+}$; **f**), macrophages and monocytes (MOMA2 $^{+}$; **g**) and cells expressing the pro- and anti-inflammatory cytokines (IL12 (h) and IL4 (i)). Mean with SEM, $n = 3$ images analysed from 3

biological replicates. Unpaired two-tailed t -test, 95% CI (b). One-way ANOVA, Tukey's test, 95% CI (c–i). **j**, Activated platelets (SELP) in 7 dpw sham and ePTFE-implanted wounds. Representative images of 3 biological replicates. **k**, Mean grey value of SELP signal. Data are mean \pm s.e.m.; $n = 3$ images analysed from 3 biological replicates. Two-tailed Student's t -test, 95% CI. Dotted lines delimit the wound area. Scale bars, 200 μ m (main images) and 100 μ m (magnified insets).



Extended Data Fig. 9 | EPFs steer matrix in vitro independently of proliferation. **a**, *En1^{cre};R26^{DTTR}* biopsies at day 0 and 6 after short treatment with DT or vehicle, immunolabelled for collagen I, III and VI. Representative images of three replicates. **b**, Collagens density. Data are mean \pm s.e.m.; $n = 3$ images analysed from 3 biological replicates. Two-way ANOVA, multiple comparison Tukey's test, 95% CI. **c**, Cell density. Data are mean \pm s.e.m.; $n = 3$ images analysed from 3 biological replicates. One-way ANOVA, multiple comparison Tukey's test, 95% CI. **d**, **e**, Time-lapse images (**d**) and contraction rate (**e**) of *En1^{cre};R26^{DTTR}* neonate fascia biopsy in culture treated with DT for 1h. Representative samples from three replicates. Contraction values obtained

from Supplementary Videos 4, 6. **f**, Fascia biopsies treated with etoposide and immunolabelled for MKI67. **g**, Fraction of MKI67⁺ cells. Data are mean \pm s.e.m.; $n = 3$ images analysed from 3 biological replicates. One-way ANOVA, Dunnett's multiple comparisons, 95% CI. **h**, **i**, Time-lapse images (**h**) and contraction rate (**i**) of neonate fascia biopsy in culture treated with 100 μ M etoposide. Representative samples from three replicates. Contraction values obtained from Supplementary Videos 4, 7. **j**, Mean (\pm s.e.m.) matrix contraction velocity during the first 25 h of imaging; $n = 25$ values from Supplementary Video 7. Two-tailed Student's *t*-test, 95% CI. Lines show the distance between two tracked points in the SHG channel. Scale bars, 50 μ m (**f**), 200 μ m (**a**, **h**) and 500 μ m (**d**).



Extended Data Fig. 10 | Fascia matrix steering precedes proliferation in vivo.

a. In situ fascia matrix labelling and EdU pulses. **b.** EdU detection in sections at defined time points. **c.** Fraction of EdU⁺ cells in the wound from total EdU⁺ cells. Data are mean \pm s.e.m.; $n=3$ (uninjured), 4 (3 dpw), 6 (7 dpw) and 4 (14 dpw)

images analysed from 3 biological replicates. One-way ANOVA, Tukey's multiple comparisons. Arrows indicate EdU-positive nuclei. Arrowheads indicate the original injury site. Broken and dotted lines delimit fascia and wounds respectively. Scale bars, 200 μ m.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Leica Application Suite v4.8
AxioVision (Carl Zeiss)
Zen (Carl Zeiss)
BD FACSDiva Software v6.1.3

Data analysis

ImageJ v1.47
GraphPad Prism v6.0
Imaris v9.1.0
Summit v4.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All other data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine sample size. Required experimental sample sizes were estimated based on previous established protocols in the field. The sample sizes were adequate as the differences between experimental groups were reproducible. All n values are clearly indicated within the figure legends.
Data exclusions	No data was excluded from the analysis.
Replication	All experiments were performed at least three times with similar results. The live imaging data showed in Fig. 3d-e and extended data figure 9d-e and h-k were generated from one representative video respectively.
Randomization	Mice were randomly divided into treatment groups.
Blinding	No experiments presented in this study required blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Primary antibodies used in IHC:

goat-anti- α SMA (Abcam, ab5694, 1:50)
 rabbit-anti-TUBB3 (Abcam, ab18207, 1:100)
 rat-anti-THY1(CD90) (Abcam, ab3105, 1:100)
 rat-anti-CD24 (BD biosciences, 553146, 1:50)
 rabbit-anti-DPP4(CD26) (Abcam, ab28340, 1:150)
 rabbit-anti-PECAM1(CD31) (Abcam, ab28364, 1:10)
 rat-anti-CD34 (Abcam, ab8158, 1:100)
 rabbit-anti-COLLAGEN I (Rockland, 600-401-103-0.1, 1:150)
 rabbit-anti-COLLAGEN III (Abcam, ab7778, 1:150)
 rabbit-anti-COLLAGEN VI (Abcam, ab6588, 1:150)
 rabbit-anti-DLK1 (Abcam, ab21682, 1:200)
 rat-anti-ERTR7 (Abcam, ab51824, 1:200)
 rat-anti-F4/80 (Abcam, ab90247, 1:400)
 rabbit-anti-LYVE1 (Abcam, ab14917, 1:100)
 rat-anti-MOMA2 (1:100, ab33451, Abcam)
 goat-anti-PDGFR α (R&D systems, AF1062, 1:50)
 rat-anti-LY6A (Biolegend, 122502, 1:150)
 rat-anti-CD44 (Abcam, ab119348, 1:100)
 rabbit-anti-NOV/CCN3 (Elabscience, E-AB-15124, 1:20)
 sheep-anti-FAP (R&D systems, AF3715, 1:100)

Secondary antibodies:

PacificBlue-, AlexaFluor488-, AlexaFluor568, or AlexaFluor647-conjugated secondary antibodies against suitable species (Life technologies, 1:500)

Antibodies used for flow cytometry:
 anti-DLK1 (Abcam, ab21682, 1:200)
 FITC-anti-CD9 (eBioscience, 11-0091-81, 1:200)
 anti-CD271(LNGFR) (Miltenyi, 130-110, 1:100)
 APC-anti-F4/80 (eBioscience, 17-4801-82, 1:200)
 AlexaFluor790-anti-NG2 (Santa Cruz, sc-166251 AF790, 1:100)
 FITC-anti-DPP4(CD26) (BioLegend, 137806, 1:200)
 PerCP-eFluor710-anti-ITGB1(CD29) (eBioscience, 46-0291-80, 1:200)
 FITC anti-CD34 (eBioscience, 11-0341-82, 1:200)
 PerCP-Cy5.5-anti-CD24 (eBioscience, 45-0242-80, 1:200)
 APC-anti-ITGA7 (Miltenyi Biotec, 130-103-356, 1:100)
 PerCP-Cy5.5-anti-LY6A(Sca1) (eBioscience, 45-5981-80, 1:200)
 PE-Vio770-anti-PDGFRA (Miltenyi, 130-105-117, 1:100)
 PerCP-Vio700-anti-CD146 (Miltenyi, 130-103-865, 1:100)
 APC-anti-PECAM1(CD31) (BioLegend, 102410, 1:200)
 eFluor660-anti-LYVE1 (eBioscience, 50-0443-82, 1:200)
 APC-LY76(TER119) (BioLegend, 116212, 1:200)
 APC-anti-EPCAM(CD326) (BioLegend, 118214, 1:200)
 APC-anti-PTPRC(CD45) (BioLegend, 103112, 1:200)

Validation

See manufacturers' notes. Antibodies were additionally validated using respective isotype antibodies in immunofluorescence assays.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Following mouse strains were used, both males and females, adult at 8-12 weeks old or neonatal at postnatal P0-P2:
 C57BL/6J wild type
 En1Cre (En1tm2(cre)Wrt)
 R26mTmG (Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo)
 R26VT2/GK3 aka "Rainbow" (Gt(ROSA)26Sortm1(CAG-EGFP,-mCerulean,-mOrange,-mCherry)llw)
 R26iDTR (Gt(ROSA)26Sortm1(HBEGF)Awai)
 RAG2-/- (Rag2tm1.1Cgn)
 Fox Chase SCID (CB17/lcr-Prkdcscid/lcrIcoCrl)

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve samples collected from field.

Ethics oversight

Government of Upper Bavaria , Germany

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

healthy donors between 18 - 65 years of age, both genders, underwent surgeries for skin tumors.

Recruitment

Fresh human skin and scar biopsies, from various anatomic locations, were collected through the Department of Dermatology and Allergology, Klinikum rechts der Isar, Technical University Munich. Informed consent was obtained from all subjects prior to skin biopsies.

Ethics oversight

Ethikkommission der Bayerischen Landesärztekammer, reference number 85/18S.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Fascia and dermis were physically separated from the back-skin of C57BL6/J or En1Cre;R26mTmG mice under the fluorescence stereomicroscope as before. Harvested tissue was minced with surgical scissors and digested with an enzymatic cocktail containing 1 mg/ml Collagenase IV, 0.5 mg/ml Hyaluronidase, and 25 U/ml DNase I (Sigma-Aldrich) at 37°C for 30 min. The resulted single cell suspension was filtered and incubated with conjugated/unconjugated primary antibodies (dilution 1:200) at 4° C for 30 min, followed by an incubation with a suitable secondary antibody when needed at 4° C for 30 min. Cells were washed and stained with Sytox blue dye (dilution 1:1000. Life technologies, S34857) for dead cell exclusion.

Instrument

Cells were subjected to flow cytometric analysis using a FACSAria III (BD Bioscience)

Software

Data was collected using the BD FACSDIVA software (BD Bioscience) and analyzed using Summit software (Cytomation)

Cell population abundance

The purity of sorted cells were determined by flow cytometric analysis of the sorted cells with the same gating strategy as during sorting.

Gating strategy

The single cells were gated base on FSC/SSC and FSC-Area/FSC-width. The viable cells were gated within SytoxBlue negative gate, and fibroblasts were gated within lineage-negative (APC-negative) gate (antibodies against lineage markers are APC-conjugated anti-CD45, anti-Ter119, anti-EpCAM, anti-CD31, anti-Lyve1, and anti-Tie2).

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

c-Jun overexpression in CAR T cells induces exhaustion resistance

<https://doi.org/10.1038/s41586-019-1805-z>

Received: 7 September 2018

Accepted: 17 October 2019

Published online: 4 December 2019

Rachel C. Lynn^{1,11}, Evan W. Weber^{1,13}, Elena Sotillo^{1,13}, David Gennert², Peng Xu¹, Zinaida Good^{1,3,4}, Hima Anbunathan¹, John Lattin¹, Robert Jones¹, Victor Tieu¹, Surya Nagaraja⁵, Jeffrey Granja², Charles F. A. de Bourcy^{6,12}, Robbie Majzner⁷, Ansuman T. Satpathy^{2,4}, Stephen R. Quake^{6,8}, Michelle Monje^{1,5,7}, Howard Y. Chang^{2,4,9} & Crystal L. Mackall^{1,4,7,10*}

Chimeric antigen receptor (CAR) T cells mediate anti-tumour effects in a small subset of patients with cancer^{1–3}, but dysfunction due to T cell exhaustion is an important barrier to progress^{4–6}. To investigate the biology of exhaustion in human T cells expressing CAR receptors, we used a model system with a tonically signaling CAR, which induces hallmark features of exhaustion⁶. Exhaustion was associated with a profound defect in the production of IL-2, along with increased chromatin accessibility of AP-1 transcription factor motifs and overexpression of the bZIP and IRF transcription factors that have been implicated in mediating dysfunction in exhausted T cells^{7–10}. Here we show that CAR T cells engineered to overexpress the canonical AP-1 factor c-Jun have enhanced expansion potential, increased functional capacity, diminished terminal differentiation and improved anti-tumour potency in five different mouse tumour models in vivo. We conclude that a functional deficiency in c-Jun mediates dysfunction in exhausted human T cells, and that engineering CAR T cells to overexpress c-Jun renders them resistant to exhaustion, thereby addressing a major barrier to progress for this emerging class of therapeutic agents.

CAR-expressing T cells demonstrate impressive response rates in B cell malignancies, but fewer than 50% of patients experience long-term disease control^{11,12} and CAR T cells have not mediated sustained responses in solid tumours³. Several factors limit the efficacy of CAR T cells, including a requirement for high antigen density for optimal CAR function enabling rapid selection of antigen loss or antigen low variants^{12–14}, the suppressive tumour microenvironment¹⁵ and intrinsic T cell dysfunction due to T cell exhaustion^{6,11,16}. T cell exhaustion has been increasingly incriminated as a cause of CAR T cell dysfunction^{6,11,16,17}, raising the prospect that engineering exhaustion-resistant CAR T cells could improve clinical outcomes.

T cell exhaustion is characterized by high expression of inhibitory receptors and widespread transcriptional and epigenetic alterations^{4,5,7,18,19}, but the mechanisms responsible for impaired function in exhausted T cells are unknown. Blockade of PD-1 can reinvigorate some exhausted T cells²⁰ but does not restore function fully, and trials using PD-1 blockade in combination with CAR T cells have not demonstrated efficacy²¹. Using a model in which healthy T cells are driven to exhaustion by the expression of a tonically signalling CAR, exhausted human T cells demonstrated widespread epigenomic dysregulation of AP-1 transcription factor-binding motifs and increased expression of the bZIP and IRF transcription factors that have been implicated in the regulation of exhaustion-related genes. Therefore, we tested the

hypothesis that dysfunction in this setting resulted from an imbalance between activating and immunoregulatory AP-1–IRF complexes by inducing overexpression of c-Jun—an AP-1 family transcription factor associated with productive T cell activation. Consistent with this hypothesis, overexpression of c-Jun rendered CAR T cells resistant to exhaustion, as demonstrated by enhanced expansion potential in vitro and in vivo, increased functional capacity, diminished terminal differentiation and improved anti-tumour potency in multiple in vivo models.

HA-28z CAR rapidly induces T cell exhaustion

Exhaustion in human T cells was recently demonstrated after expression of a CAR incorporating the disialoganglioside (GD2)-specific 14g2a scFv, CD3ζ and CD28 signalling domains (GD2-28z), as a result of tonic signalling mediated via antigen-independent aggregation⁶. Here we show that CARs incorporating the 14g2a-E101K scFv, which demonstrate higher affinity for GD2²² (HA-28z), display a more severe exhaustion phenotype (Extended Data Fig. 1a–c). In contrast to CD19-28z CAR T cells (without tonic signalling), HA-28z CAR T cells develop profound features of exhaustion, including reduced expansion in culture, increased expression of inhibitory receptors, exaggerated effector differentiation, and diminished IFNγ and markedly decreased IL-2 production after stimulation (Fig. 1a–d, Extended Data Fig. 1d, e). The

¹Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ²Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA. ³Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. ⁴Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA. ⁵Department of Neurology, Stanford University, Stanford, CA, USA. ⁶Departments of Bioengineering and Applied Physics, Stanford University, Stanford, CA, USA. ⁷Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA. ⁸Chan Zuckerberg Biohub, San Francisco, CA, USA. ⁹Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. ¹⁰Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ¹¹Present address: Lyell Immunopharma, South San Francisco, CA, USA. ¹²Present address: Chan Zuckerberg Initiative, San Francisco, CA, USA. ¹³These authors contributed equally: Evan W. Weber, Elena Sotillo. *e-mail: cmackall@stanford.edu

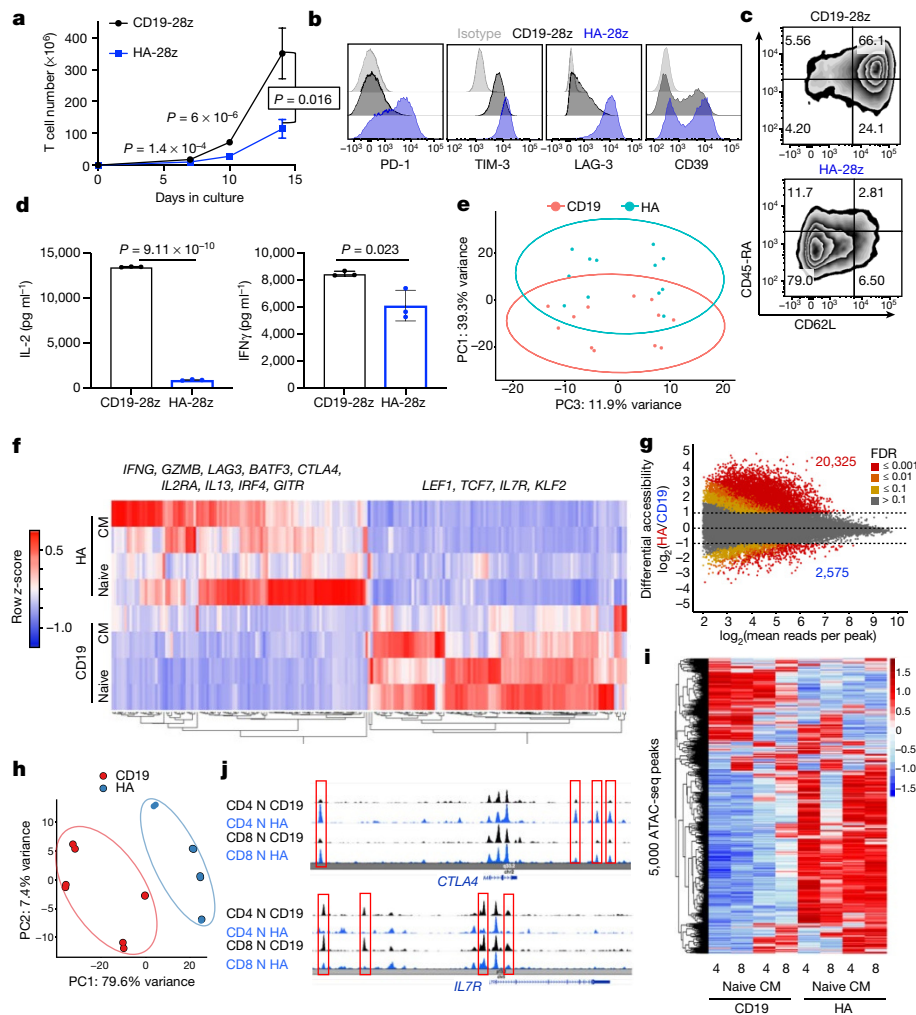


Fig. 1 | HA-28z CAR T cells manifest phenotypic, functional, transcriptional and epigenetic hallmarks of T cell exhaustion. **a**, Primary T cell expansion. Data are mean \pm s.e.m. from $n = 10$ independent experiments. **b**, Surface expression of exhaustion-associated markers. **c**, Surface expression of CD45RA and CD62L to distinguish T memory stem cells (CD45RA⁺CD62L⁺), central memory cells (CD45RA⁺CD62L⁻), and effector memory cells (CD45RA⁺CD62L⁻). **d**, IL-2 (left) and IFN γ (right) release after 24-h co-culture with CD19⁺GD2⁺ Nalm6-GD2 leukaemia cells. Data are mean \pm s.d. from triplicate wells. In **b–d**, one representative donor (of $n = 10$ experiments) is shown for each assay. *P* values determined by unpaired two-tailed *t*-tests. **e**, Principal component analysis (PCA) of global transcriptional profiles of naive- and central-memory-derived CD19-28z (CD19) or HA-28z (HA) CAR T cells at days 7, 10 and 14 in

culture. PC1 (39.3% variance) separates CD19-28z from HA-28z CAR T cells. **f**, Gene expression of the top 200 genes driving PC1. Genes of interest in each cluster are listed. **g**, Differentially accessible chromatin regions (peaks) in CD8⁺ CD19-28z and HA-28z CAR T cells. Both naive and central memory cell subsets are incorporated for each CAR. **h**, PCA of ATAC-seq chromatin accessibility in CD19-28z or HA-28z CAR T cells. PC1 (76.9% variance) separates CD19-28z from HA-28z CAR samples. **i**, Global chromatin accessibility profile of CD4⁺ and CD8⁺ CD19-28z and HA-28z CAR T cells derived from naive (N) and central memory (CM) subsets. Top 5,000 peaks. **j**, Differentially accessible enhancer regions in CD19- and HA-28z CAR T cells in the *CTLA4* (top) or *IL7R* (bottom) loci. Unless noted otherwise, all analyses were done on day 10 of culture. *GITR* is also known as *TNFRSF18*.

functional defects are due to exhaustion-associated dysfunction rather than suboptimal interaction of the HA-28z CAR with its target GD2, because they are also observed in CD19-28z CAR T cells when HA-28z CAR is co-expressed using a bi-cistronic vector (Extended Data Fig. 1f). Principal component analysis (PCA) of RNA-sequencing (RNA-seq) data demonstrated that the strongest driver of transcriptional variance was the presence of the exhausting HA-28z versus control CD19-28z CAR (Fig. 1e), although some cell-type-specific differences were observed (Extended Data Fig. 1g).

The top 200 most differentially expressed genes (Fig. 1f, Supplementary Table 1) included activation-associated genes (*IFNG*, *GZMB* and *IL2RA*), inhibitory receptors (*LAG3* and *CTLA4*) and inflammatory chemokines or cytokines (*CXCL8*, *IL13* and *IL1A*), and genes associated with naive and memory T cells (*IL7R*, *TCF7*, *LEF1* and *KLF2*), which overlapped with gene sets described in chronic lymphocytic choriomeningitis virus (LCMV) mouse models⁴ (Extended Data Fig. 1h). Single-cell

RNA-seq analysis of GD2-28z versus CD19-28z CAR T cells revealed similar differential gene expression as HA-28z CAR T cells (Extended Data Fig. 2).

T cell exhaustion is associated with widespread epigenetic changes^{18,20}. Using ATAC-seq (assay for transposase-accessible chromatin using sequencing) analysis²³ (Fig. 1g, Extended Data Figs. 3, 4a), we observed that CD8⁺ HA-28z CAR T cells displayed more than 20,000 unique differentially accessible chromatin regions (peaks) compared with less than 3,000 unique peaks in CD8⁺ CD19-28z CAR T cells (false discovery rate (FDR) < 0.1 and \log_2 -transformed fold change > 1). Principal component analysis (PCA) revealed HA-28z versus CD19-28z CAR as the strongest driver of differential chromatin states (PC1 variance 79.6%, Fig. 1h), with weaker but observable differences observed between naive and central memory cells (PC2 variance 7.4%), and CD4 versus CD8 subsets (PC3 variance 6.5%; Extended Data Fig. 4b). Clustering the top 5,000 differentially accessible regions revealed a similar epigenetic

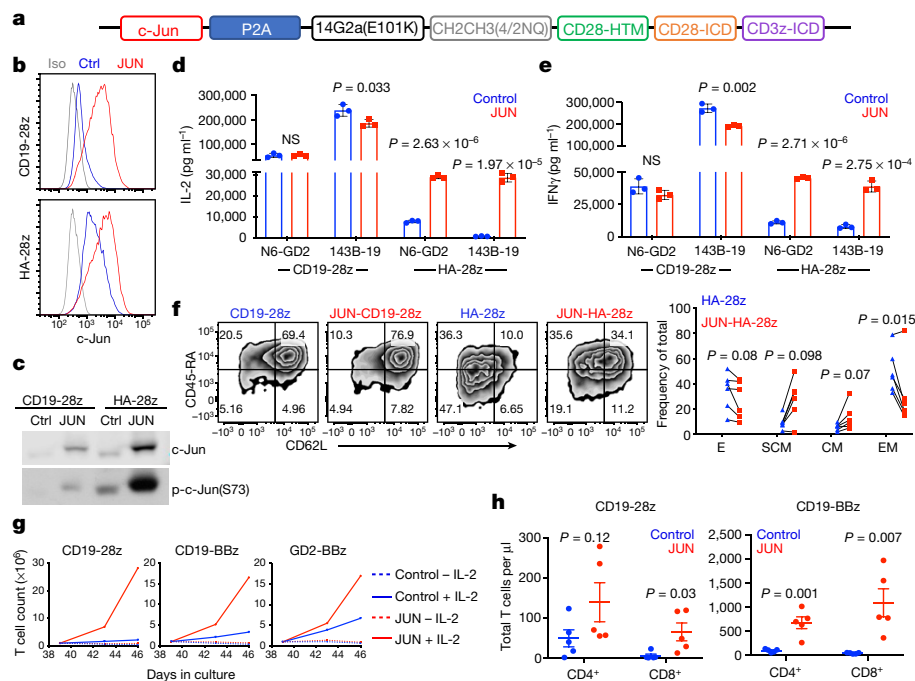


Fig. 3 | c-Jun overexpression enhances the function of exhausted CAR T cells. **a**, JUN-P2A-HA-CAR expression vector. HTM, hinge/transmembrane; ICD, intracellular domain. **b**, Intracellular c-Jun expression in control (Ctrl) and JUN CAR T cells at day 10 by flow cytometry. Grey denotes isotype control. **c**, Immunoblot for total c-Jun and phosphorylated c-Jun (p-c-Jun(S73)) in control (Ctrl) and JUN CAR T cells at day 10. **d, e**, IL-2 (**d**) and IFN γ (**e**) production after 24-h co-culture of control or JUN CD19-28z and HA-28z CAR T cells in response to antigen-positive tumour cells. Data are mean \pm s.d. of triplicate wells. *P* values determined by unpaired two-tailed *t*-tests. One representative donor. Fold change across $n = 8$ donors in Extended Data Fig. 5. **f**, Left, flow cytometry showing representative expression of CD45RA and CD62L in control and JUN CAR T cells at day 10. Right, relative frequency of effector

(E; CD45RA⁺CD62L⁻), stem-cell memory (SCM; CD45RA⁺CD62L⁺), central memory (CM; CD45RA⁻CD62L⁺), and effector memory (EM; CD45RA⁻CD62L⁻) in CD8⁺ control or JUN-HA-28z CAR T cells ($n = 6$ donors from independent experiments). Lines indicate paired samples from the same donor. *P* values determined by paired two-tailed *t*-tests. **g**, On day 39, 1×10^6 viable T cells from Extended Data Fig. 5f were re-plated and cultured for 7 days with or without IL-2. **h**, Control or JUN CD19-28z or CD19-BBz CAR T cells from **g** were cryopreserved on day 10 and later thawed, rested overnight in IL-2 and 5×10^6 cells were then injected intravenously into healthy NSG mice. On day 25 after infusion, peripheral blood T cells were quantified by flow cytometry. Data are mean \pm s.e.m. of $n = 5$ mice per group. *P* values determined by unpaired two-tailed *t*-tests.

Data Fig. 5d, e). Together, the data are consistent with a model in which c-Jun overexpression is functionally more significant in exhausted T cells, which express higher levels of immunomodulatory bZIP and IRF transcription factors.

Overexpression of c-Jun also enhanced long-term proliferative capacity, which is associated with anti-tumour effects in solid tumours³¹, in CAR T cells without tonic signalling (CD19-28z, CD19-BBz) (Extended Data Fig. 5f). Enhanced proliferation remained IL-2-dependent, as expansion immediately ceased after IL-2 withdrawal (Fig. 3g, Extended Data Fig. 5g). Expanding CD8⁺ JUN-CD19-28z CAR T cells displayed diminished exhaustion markers and an increased frequency of cells bearing the stem-cell memory phenotype compared with control CD19-28z CAR T cells (Extended Data Fig. 5h–j). c-Jun overexpression also increased homeostatic expansion of both CD19-28z and CD19-BBz CAR T cells in tumour-free NSG (NOD-SCID *IL2rg*-null) mice (Fig. 3h), which led to accelerated GVHD in the JUN-CD19-BBz CAR-T-cell-treated mice. Together, the data demonstrate that c-Jun overexpression mitigates T cell exhaustion in numerous CARs tested, including those incorporating CD28 or 4-1BB costimulatory domains, and regardless of whether exhaustion is driven by long-term expansion or tonic signalling.

Molecular mechanisms of c-Jun in exhaustion

To explore the mechanism by which c-Jun overexpression prevents T cell dysfunction, we compared ATAC-seq and RNA-seq results of HA-28z and JUN-HA-28z CAR T cells. Overexpression of c-Jun did not change

the epigenetic profile but substantially modulated the transcriptome, with 319 genes differentially expressed in JUN and control HA-28z CAR T cells (Extended Data Fig. 6a–c), including reduced expression of exhaustion-associated genes and increased expression of memory genes. Using DAVID, we confirmed that genes changed by c-Jun are highly enriched for AP-1 family binding sites (Extended Data Fig. 6d), which suggests that gene expression changes were mediated by AP-1 family transcription factors.

We postulated that c-Jun overexpression could rescue exhausted T cells by direct transcriptional activation of AP-1 target genes and/or by indirectly disrupting immunoregulatory AP-1–IRF transcriptional complexes^{7,9} that drive exhaustion-associated gene expression (AP-1i) (Extended Data Fig. 6e). To test these non-mutually-exclusive hypotheses, we first evaluated a panel of c-Jun mutants predicted to be deficient in transcriptional activation (JUN-AA, JUN- Δ 8, JUN- Δ TAD), DNA binding (JUN- Δ basic) or dimerization (JUN- Δ Leu, JUN- Δ bZIP)^{32–34} (Fig. 4a, Extended Data Fig. 6f). JUN-AA and JUN- Δ 8 both equivalently increased IL-2 and IFN γ production compared to wild-type c-Jun in HA-28z CAR T cells (Fig. 4b), whereas JUN- Δ TAD demonstrated partial rescue in IL-2 production. Conversely, C-terminal mutants (JUN- Δ basic, JUN- Δ Leu and JUN- Δ bZIP), which were unable to bind chromatin (Extended Data Fig. 6g), did not rescue cytokine production in exhausted HA-28z CAR T cells (Fig. 4b). Furthermore, c-Jun overexpression substantially decreased levels of AP-1i, as evidenced by diminished mRNA levels (Extended Data Fig. 6h), reduced total and chromatin-bound JunB, BATF and BATF3 proteins, and reduced JunB–BATF complexes (Extended Data

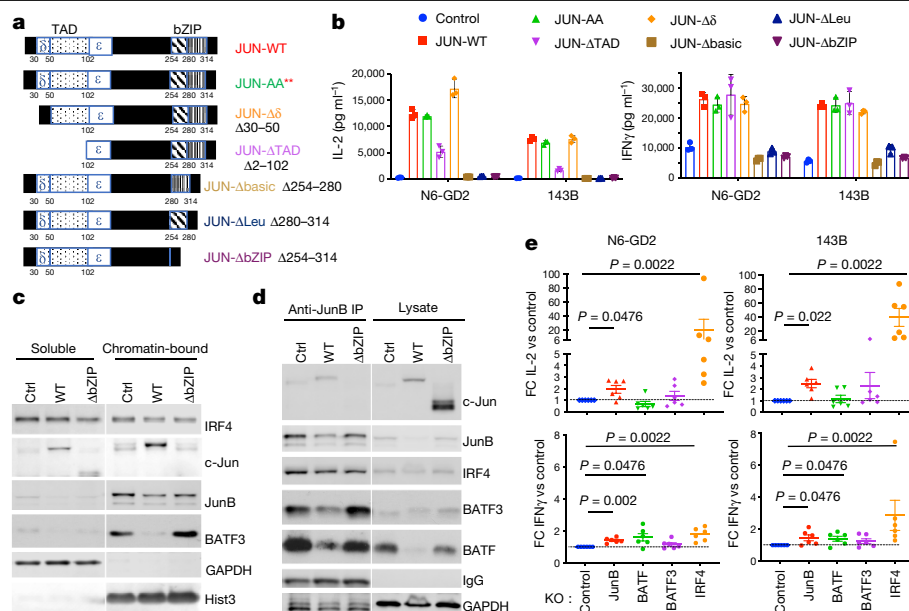


Fig. 4 | c-Jun functional rescue of exhaustion requires bZIP dimerization but is independent of transactivation. a, Schematic of c-Jun protein showing N-terminal transactivation domain (TAD) and C-terminal bZIP domain deletion mutants. Red asterisks denote JNP sites at Ser63 and Ser73 mutated to alanine in JUN-AA. **b**, IL-2 (left) and IFNγ (right) production by control or JUN-HA-28z CAR T cells expressing the indicated c-Jun variant after 24-h stimulation with Nalm6-GD2 (N6-GD2) or 143B target cells. Data are mean ± s.d. of triplicate wells; representative of three independent experiments. **c**, Immunoblot of indicated AP-1 and IRF proteins in control, JUN-WT or JUN-ΔbZIP HA-28z CAR

T cells in soluble or chromatin-bound lysis fractions. **d**, Immunoblot of indicated AP-1 and IRF proteins in control, JUN-WT or JUN-ΔbZIP HA-28z CAR T cells in total lysate (right) or after JunB immunoprecipitation (IP, left). **e**, Fold change (FC) in IL-2 (top) and IFNγ (bottom) production in AP-1 or IRF4 CRISPR-knockout (KO) HA-28z CAR T cells after 24-h stimulation with Nalm6-GD2 or 143B target cells. Fold change in cytokine production is normalized to control HA-28z CAR T cells. Data are mean ± s.e.m. of *n* = 6 independent experiments. *P* values determined using nonparametric Mann-Whitney *U* tests.

Fig. 7a–c). Importantly, c-Jun-mediated displacement of JunB, BATF and BATF3 from chromatin and reduced JunB–BATF complexes were dependent on the ability of c-Jun to partner with AP-1 family members (Fig. 4c, d). Consistent with this, c-Jun and IRF4 chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) analysis identified no novel c-Jun-binding sites after c-Jun overexpression. Instead, the vast majority of sites bound by c-Jun are also bound by IRF4 (and probably BATF), consistent with c-Jun overexpression increasing binding almost exclusively at AP-1–IRF composite elements, including near exhaustion-associated genes regulated by IRF4, and genes associated with increased T cell proliferation and functional activation (Extended Data Fig. 7d–h). Finally, JunB-knockout, BATF-knockout and especially IRF4-knockout significantly increased IL-2 and IFNγ production in HA-28z CAR T cells (Fig. 4e, Extended Data Fig. 7i, j). Time-course experiments using a drug regulatable expression model of c-Jun revealed that full rescue required c-Jun overexpression during both T cell expansion and antigen stimulation (Extended Data Fig. 8), consistent with a model in which c-Jun overexpression both modulates molecular reprogramming during the development of exhaustion and augments responses during acute stimulation downstream of antigen encounter. Together, the data are consistent with a model in which an overabundance of AP-1–IRF complexes drives the exhaustion transcriptional program and c-Jun overexpression prevents exhaustion by decreasing and/or displacing AP-1 complexes from chromatin.

JUN CAR T cells enhance anti-tumour activity in vivo

Using a Nalm6-GD2⁺ leukaemia model, we confirmed functional superiority of JUN-HA-28z CAR T cells in vivo (Fig. 5a–c), which required c-Jun dimerization but not transactivation (Extended Data Fig. 7k, l). In an in vitro model of limiting antigen dilution, JUN-HA-28z CAR T cells produced greater maximal IL-2 and IFNγ and manifested a lower threshold for antigen-induced IL-2 secretion (Fig. 5d, e). Limiting target antigen

expression is increasingly recognized to limit CAR functionality as observed after treatment of CD22-BBz-CAR T cells in patients with relapsed or refractory leukaemia^{12,13,35}. We therefore assessed whether c-Jun overexpression could enhance the capacity to target antigen-low tumour cells. In response to CD22^{low} leukaemia, JUN-CD22-BBz CAR T cells exhibited increased cytokine production in vitro (Fig. 5f–h) and markedly increased anti-tumour activity in vivo (Fig. 5i–l) compared with control CD22-BBz CAR T cells. Similar results were observed in a CD19^{low} Nalm6 leukaemia model (Extended Data Fig. 9a–f).

c-Jun decreases hypofunction within solid tumours

c-Jun overexpression also enhanced the functionality of CARs targeting solid tumours. JUN-Her2-BBz CAR T cells prevented 143B osteosarcoma tumour growth in vivo, markedly improved long-term survival, and greatly increased T cell expansion (Extended Data Fig. 9g–i). Similar results were observed when comparing GD2-BBz and JUN-GD2-BBz CAR T cells against 143B (Extended Data Fig. 9j–n). c-Jun overexpression increased the frequency of total and CAR⁺ Her2-BBz T cells within tumours (Fig. 6a, b), reduced expression of exhaustion markers PD-1 and CD39 (Fig. 6c), and substantially increased cytokine production after ex vivo re-stimulation (Fig. 6d, e, Extended Data Fig. 10a–c). Single-cell RNA-seq of purified tumour infiltrating JUN-Her2-BBz CAR T cells demonstrated increased frequency of cells within the G2/M and S phases of the cell cycle (Fig. 6f), a more activated transcriptional program (as measured by *IL2RA* and *CD38*), and downregulation of numerous exhaustion-associated genes (*PDCD1*, *BTLA*, *TIGIT*, *CD200*, *ENTPD1* and *NR4A2*) (Fig. 6g and Extended Data Fig. 10d–g). Finally, a small cluster of T cells characterized by high *IL7R* expression (*IL7R*, *KLF2*, *CD27*, *TCF7* and *SELL*) was preserved in tumours treated with JUN CAR T cells but not those receiving control Her2-BBz CAR T cells (Extended Data Fig. 10g), consistent with c-Jun-induced maintenance of a memory-like population capable of self-renewal.

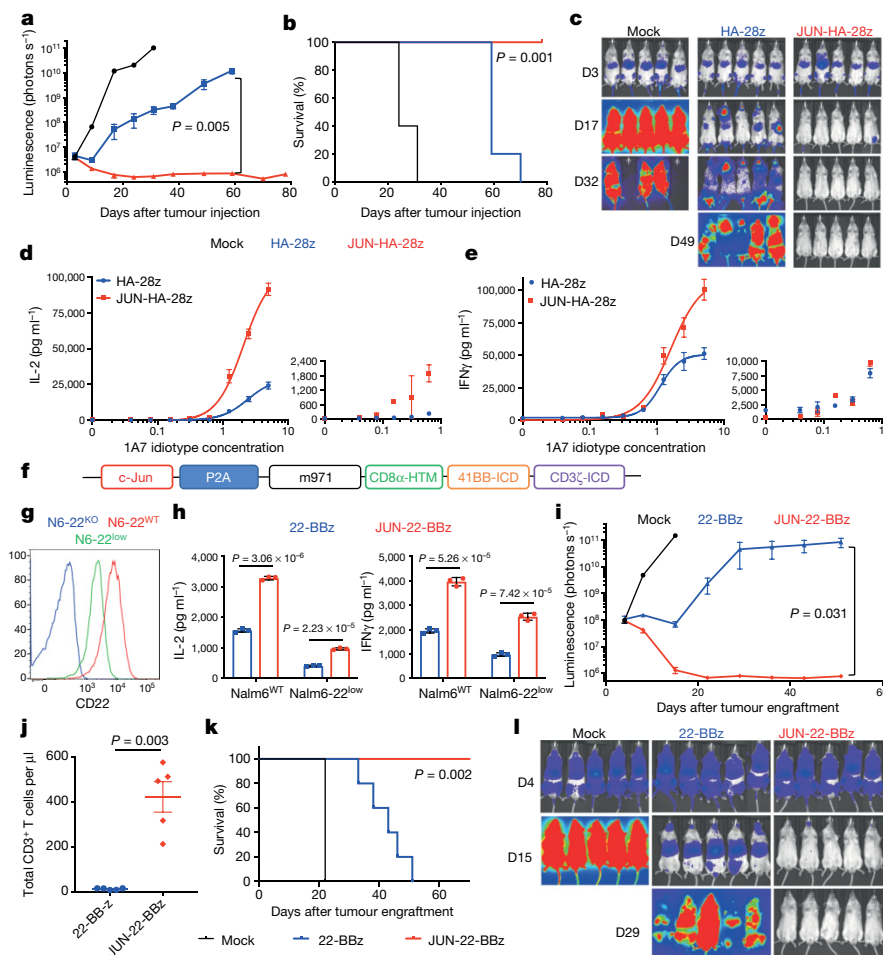


Fig. 5 | JUN-modified CAR T cells increase in vivo activity against leukaemia and enhance T cell function under suboptimal stimulation. **a–c**, NSG mice were injected intravenously with 1×10^6 Nalm6-GD2 leukaemia cells, and then 3×10^6 mock, HA-28z or JUN-HA-28z CAR⁺ T cells were given intravenously on day 3. **a, c**, Tumour progression was monitored using bioluminescent imaging. Scales are normalized for all time points. **d**, day. **b**, JUN-HA-28z CAR T cells induced long-term tumour-free survival. Data are mean \pm s.e.m. of $n = 5$ mice per group. Reproducible in three independent experiments; however, in some experiments, long-term survival was diminished owing to outgrowth of GD2(-) Nalm6 clones. **d, e**, IL-2 (**d**) and IFN γ (**e**) production after 24 h stimulation of control or JUN HA-28z CAR T cells with immobilized 1A7 anti-CAR idiotype antibody. Each curve was fit with nonlinear dose response kinetics to determine half-maximal effective concentration (EC_{50}) values. Smaller graphs (right) highlight antibody concentrations less than $1 \mu\text{g ml}^{-1}$. Data are mean \pm s.d. of triplicate wells; representative of two independent experiments.

f, JUN-CD22-BBz retroviral vector. **g**, CD22 surface expression on Nalm6 wild-type (N6-22^{WT}), Nalm6-CD22-knockout (N6-22^{KO}) and Nalm6-22^{KO} plus CD22^{low} (N6-22^{low}) cells. **h**, IL-2 (left) and IFN γ (right) release after co-culture of Nalm6 and Nalm6-22^{low} cells with control or JUN-CD22-BBz CAR T cells. Data are mean \pm s.d. of triplicate wells; representative of three independent experiments. **i–l**, NSG mice were inoculated with 1×10^6 Nalm6-22^{low} leukaemia cells intravenously. On day 4, 3×10^6 mock, control or JUN-CD22-BBz CAR⁺ T cells were transferred intravenously. **i, j**, Tumour growth was monitored by bioluminescent imaging. **j**, Mice receiving JUN-CD22-BBz CAR T cells display increased peripheral blood T cells on day 23. **k**, Long-term survival of CAR-treated mice. Data in **i** and **j** are mean \pm s.e.m. of $n = 5$ mice per group; representative of two independent experiments. Unless otherwise noted, P values determined by unpaired two-tailed t -tests. Survival curves were compared using the log-rank Mantel–Cox test.

Discussion

Several lines of evidence implicate exhaustion in limiting the potency of CAR T cells^{6,11,16,17}. Using a tonically signalling CAR that can induce the hallmark features of exhaustion in a controlled in vitro culture system, we identified AP-1-related bZIP–IRF families as major factors that drive exhaustion-associated gene expression. We tested the hypothesis that exhaustion-associated dysfunction results from increased levels of AP-1–IRF complexes leading to a functional deficiency in activating AP-1 Fos/Jun heterodimers. Consistent with this model, c-Jun overexpression prevented phenotypic and functional hallmarks of exhaustion and improved anti-tumour control in five tumour models, including the clinically relevant GD2-BBz and Her2-BBz CAR T cells and CD19 CAR T cells subjected to prolonged ex vivo expansion. JUN CAR T cells also demonstrated increased potency when encountering tumour cells with low antigen density.

Mechanistically, c-Jun overexpression could work by directly enhancing c-Jun-mediated transcriptional activation of genes such as *IL2*, and/or indirectly by disrupting or displacing AP-1i. Substantial orthogonal data are consistent with the indirect displacement model. First, the inability of Fos overexpression to enhance function is consistent with the displacement model, as Fos has not been described to heterodimerize with BATF proteins. Second, c-Jun mutant experiments demonstrated a crucial role for dimerization but not transactivation in the biology observed. Third, we observed a reduction in total and chromatin-bound JunB, BATF and BATF3 after c-Jun overexpression, and could reproduce functional enhancement of exhausted T cells after knockout of *IRF4* and *JUNB*. An indirect model in which c-Jun blocks access of AP-1i complexes to enhancer regions is also consistent with the previous finding that BACH2 protects from terminal effector differentiation by blocking AP-1 sites³⁶ as terminal effector differentiation is a hallmark of exhaustion in

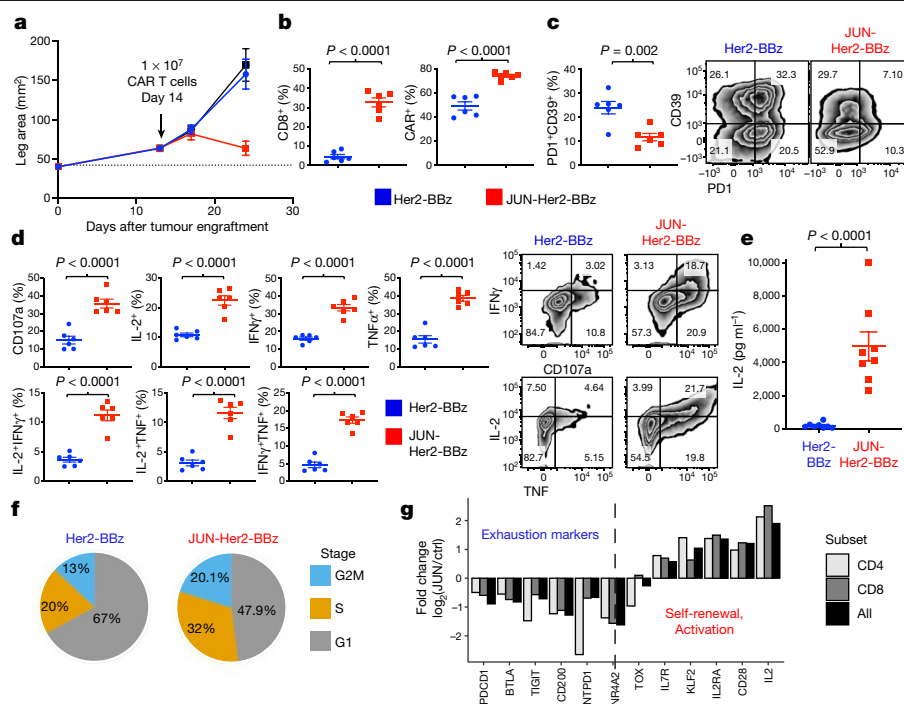


Fig. 6 | c-Jun overexpression enhances CAR T cell efficacy and decreases hypofunction within solid tumours. NSG mice were inoculated with 1×10^6 143B osteosarcoma cells via intramuscular injection, and then 1×10^7 mock, Her2-BBz or JUN-Her2-BBz CAR T cells were given intravenously on day 14. **a**, Tumour growth (monitored by caliper measurements). **b–g**, On day 28, mice were euthanized and tumour tissue was collected and mechanically dissociated. Single-cell suspensions were labelled for analysis by flow cytometry (**b, c**), re-stimulated with Nalm6-Her2⁺ target cells and analysed for intracellular cytokine production (**d**), or sorted by FACS to isolate live, human CD45⁺ tumour-infiltrating lymphocytes (TILs) (**e–g**). **b**, Left, CD8⁺ cells as a proportion of total live tumour cells. Right, CAR⁺ cells as a proportion of total

live CD8⁺ cells. **c**, PD-1⁺CD39⁺ cells as a frequency of total live CD8⁺ (left) with representative contour plots (right). **d**, Frequency of indicated cytokine- or CD107a-producing cells after 5-h re-stimulation with Nalm6-Her2⁺ target cells. Gated on total, live CD8⁺ T cells (left) with representative contour plots (right). **e**, IL-2 secretion after 24-h re-stimulation of sorted CD45⁺ TILs with Nalm6-Her2⁺ target cells. Data in **a–e** are mean \pm s.e.m. of $n = 6–8$ mice per group. Unless otherwise noted, P values determined by unpaired two-tailed t -tests. **f**, Relative frequency of sorted CD45⁺ TILs in each phase of the cell cycle as determined by single-cell RNA-seq. **g**, The log₂-transformed fold change in JUN compared with control Her2-BBz CAR T cells for the indicated transcripts.

our model and is prevented by c-Jun overexpression. Another related hypothesis suggests that exhaustion results from partner-less NFAT in the absence of AP-1³⁷ and several recent publications implicated the NFAT-driven transcription factors NR4A^{38,39} and TOX^{40–42} in T cell exhaustion. Overexpression of NR4A1 was shown to displace chromatin-bound c-Jun³⁸, suggesting that competition with NR4A family members might also contribute to the effects described here. Future studies are warranted to understand the functional overlap of NFAT, TOX and NR4A in c-Jun-overexpressing CAR T cells.

The impressive effects of c-Jun overexpression in several preclinical tumour models raise the prospect of clinical testing of JUN CAR T cells. c-Jun is the cellular homologue of the viral oncogene v-Jun⁴³, and c-Jun expression has been described in cancer^{44,45}. However, c-Jun has not been implicated as an oncogene in mature T cells, which appear to be generally resistant to transformation, and we see no evidence for transformation in these studies. *Ras*-mediated transformation in rodent models requires JNP⁴⁶, therefore, the JUN-AA mutant, which equally rescues CAR T cell function, could be implemented to mitigate theoretical oncogenic risk. Future work is necessary to determine whether c-Jun overexpression might enhance the risk of other toxicities, including on-target and off-target effects.

In summary, our findings highlight the power of a deconstructed model of human T cell exhaustion to interrogate the biology of this complex phenomenon. Using this approach, we discovered a fundamental role for the AP-1-bZIP family in human T cell exhaustion and demonstrate that overexpression of c-Jun renders CAR T cells resistant

to exhaustion, enhances their ability to control tumour growth in vivo, and improves the recognition of antigen-low targets, thus addressing major barriers to progress with this class of therapeutic agents.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1805-z>.

- Maude, S. L. et al. Tisagenlecleucel in children and young adults with b-cell lymphoblastic leukemia. *N. Engl. J. Med.* **378**, 439–448 (2018).
- Neelapu, S. S. et al. Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. *N. Engl. J. Med.* **377**, 2531–2544 (2017).
- June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S. & Milone, M. C. CAR T cell immunotherapy for human cancer. *Science* **359**, 1361–1365 (2018).
- Wherry, E. J. et al. Molecular signature of CD8⁺ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670–684 (2007).
- Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486–499 (2015).
- Long, A. H. et al. 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. *Nat. Med.* **21**, 581–590 (2015).
- Man, K. et al. Transcription factor IRF4 promotes CD8⁺ T cell exhaustion and limits the development of memory-like T cells during chronic infection. *Immunity* **47**, 1129–1141 (2017).
- Li, P. et al. BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**, 543–546 (2012).
- Murphy, T. L., Tussiwand, R. & Murphy, K. M. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat. Rev. Immunol.* **13**, 499–509 (2013).

10. Quigley, M. et al. Transcriptional analysis of HIV-specific CD8⁺ T cells shows that PD-1 inhibits T cell function by upregulating BATF. *Nat. Med.* **16**, 1147–1151 (2010).
11. Fraietta, J. A. et al. Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nat. Med.* **24**, 563–571 (2018).
12. Walker, A. J. et al. Tumor antigen and receptor densities regulate efficacy of a chimeric antigen receptor targeting anaplastic lymphoma kinase. *Mol. Ther.* **25**, 2189–2201 (2017).
13. Fry, T. J. et al. CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. *Nat. Med.* **24**, 20–28 (2018).
14. Hegde, M., Moll, A. J., Byrd, T. T., Louis, C. U. & Ahmed, N. Cellular immunotherapy for pediatric solid tumors. *Cytotherapy* **17**, 3–17 (2015).
15. Long, A. H. et al. Reduction of MDSCs with all-trans retinoic acid improves CAR therapy efficacy for sarcomas. *Cancer Immunol. Res.* **4**, 869–880 (2016).
16. Eyquem, J. et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).
17. Fraietta, J. A. et al. Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. *Nature* **558**, 307–312 (2018).
18. Sen, D. R. et al. The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).
19. Bengsch, B. et al. Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. *Immunity* **48**, 1029–1045 (2018).
20. Pauken, K. E. et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).
21. Heczey, A. et al. CAR T cells administered in combination with lymphodepletion and PD-1 inhibition to patients with neuroblastoma. *Mol. Ther.* **25**, 2214–2224 (2017).
22. Horwacik, I. et al. Structural basis of GD2 ganglioside and mimetic peptide recognition by 14G2a antibody. *Mol. Cell. Proteomics* **14**, 2577–2590 (2015).
23. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
24. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
25. Philip, M. et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature* **545**, 452–456 (2017).
26. Meixner, A., Karreth, F., Kenner, L. & Wagner, E. F. JunD regulates lymphocyte proliferation and T helper cell cytokine expression. *EMBO J.* **23**, 1325–1335 (2004).
27. Chiu, R., Angel, P. & Karin, M. Jun-B differs in its biological properties from, and is a negative regulator of, c-Jun. *Cell* **59**, 979–986 (1989).
28. Echlin, D. R., Tae, H. J., Mitin, N. & Taparowsky, E. J. B-ATF functions as a negative regulator of AP-1 mediated transcription and blocks cellular transformation by Ras and Fos. *Oncogene* **19**, 1752–1763 (2000).
29. Sade-Feldman, M. et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998–1013 (2018).
30. Dérjard, B. et al. JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. *Cell* **76**, 1025–1037 (1994).
31. D'Angelo, S. P. et al. Antitumor activity associated with prolonged persistence of adoptively transferred NY-ESO-1^{c259} T cells in synovial sarcoma. *Cancer Discov.* **8**, 944–957 (2018).
32. Adler, V., Franklin, C. C. & Kraft, A. S. Phorbol esters stimulate the phosphorylation of c-Jun but not v-Jun: regulation by the N-terminal delta domain. *Proc. Natl Acad. Sci. USA* **89**, 5341–5345 (1992).
33. Bannister, A. J., Oehler, T., Wilhelm, D., Angel, P. & Kouzarides, T. Stimulation of c-Jun activity by CBP: c-Jun residues Ser63/73 are required for CBP induced stimulation *in vivo* and CBP binding *in vitro*. *Oncogene* **11**, 2509–2514 (1995).
34. Weiss, C. et al. JNK phosphorylation relieves HDAC3-dependent suppression of the transcriptional activity of c-Jun. *EMBO J.* **22**, 3686–3695 (2003).
35. Majzner, R. G. & Mackall, C. L. Tumor antigen escape from CAR T-cell therapy. *Cancer Discov.* **8**, 1219–1226 (2018).
36. Roychoudhuri, R. et al. BACH2 regulates CD8⁺ T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat. Immunol.* **17**, 851–860 (2016).
37. Martinez, G. J. et al. The transcription factor NFAT promotes exhaustion of activated CD8⁺ T cells. *Immunity* **42**, 265–278 (2015).
38. Liu, X. et al. Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature* **567**, 525–529 (2019).
39. Chen, J. et al. NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534 (2019).
40. Seo, H. et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8⁺ T cell exhaustion. *Proc. Natl Acad. Sci. USA* **116**, 12410–12415 (2019).
41. Alfei, F. et al. TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature* **571**, 265–269 (2019).
42. Khan, O. et al. TOX transcriptionally and epigenetically programs CD8⁺ T cell exhaustion. *Nature* **571**, 211–218 (2019).
43. Bohmann, D. et al. Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* **238**, 1386–1392 (1987).
44. Mariani, O. et al. JUN oncogene amplification and overexpression block adipocytic differentiation in highly aggressive sarcomas. *Cancer Cell* **11**, 361–374 (2007).
45. Shaulian, E. AP-1—The Jun proteins: Oncogenes or tumor suppressors in disguise? *Cell. Signal.* **22**, 894–899 (2010).
46. Behrens, A., Jochum, W., Sibilio, M. & Wagner, E. F. Oncogenic transformation by ras and fos is mediated by c-Jun N-terminal phosphorylation. *Oncogene* **19**, 2657–2663 (2000).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Viral vector construction

MSGV retroviral vectors encoding the following CARs were previously described: CD19-28z, CD19-BBz, GD2-28z, GD2-BBz, Her2-BBz and CD22-BBz. To create the HA-28z CAR, a point mutation was introduced into the 14G2a scFv of the GD2-28z CAR plasmid to create the E101K mutation. The '4/2NQ' mutations⁴⁷ were introduced into the CH2CH3 domains of the IgG1 spacer region to diminish Fc receptor recognition for in vivo use of HA-28z CAR T cells.

Codon-optimized cDNAs encoding c-Jun (*JUN*), c-Fos (*FOS*), and truncated NGFR (tNGFR; *NGFR*) were synthesized by IDT and cloned into lentiviral expression vectors to create JUN-P2A-FOS, and JUN and FOS single expression vectors co-expressing tNGFR under the separate PGK promoter. JUN-P2A was then subcloned into the XhoI site of MSGV CAR vectors using the In-Fusion HD cloning kit (Takara) upstream of the CAR leader sequence to create JUN-P2A-CAR retroviral vectors. For JUN-AA, point mutations were introduced to convert Ser63 and Ser73 to Ala. The other JUN mutants were cloned to remove portions of the protein as described in Fig. 4. The *Escherichia coli* DHFR-destabilization domain (DD) sequence was inserted upstream of Jun to create JUN-DD fusion constructs. In some cases, GFP cDNA was subcloned upstream of the CAR to create GFP-P2A-CAR vector controls. For bi-cistronic CAR retroviral vectors, the HA-28z or Her2-28z CAR was cloned downstream of a codon-optimized CD19-28z CAR to create CD19-28z-P2A-HA-28z and CD19-28z-P2A-Her2-28z dual CAR expression vectors.

Viral vector production

Retroviral supernatant was produced in the 293GP packaging cell line as previously described⁶. In brief, 70% confluent 293GP 20-cm plates were co-transfected with 20 µg MSGV vector plasmid and 10 µg RD114 envelope plasmid DNA using Lipofectamine 2000. Medium was replaced at 24 and 48 h after transfection. The 48-h and 72-h viral supernatants were collected, centrifuged to remove cell debris, and frozen at -80 °C for future use. Third-generation, self-inactivating lentiviral supernatant was produced in the 293T packaging cell line. In brief, 70% confluent 293T 20-cm plates were co-transfected with 18 µg pELNS vector plasmid, and 18 µg pRSV-Rev, 18 µg pMDLg/pRRE (Gag/Pol) and 7 µg pMD2.G (VSVG envelope) packaging plasmid DNA using Lipofectamine 2000. Medium was replaced at 24 h after transfection. The 24-h and 48-h viral supernatants were collected, combined and concentrated by ultracentrifugation at 28,000 rpm for 2.5 h. Concentrated lentiviral stocks were frozen at -80 °C for future use.

T cell isolation

Healthy donor buffy coats were collected by and purchased from the Stanford Blood Center under an IRB-exempt protocol. Primary human T cells were isolated using the RosetteSep Human T cell Enrichment kit (Stem Cell Technologies) according to the manufacturer's protocol using Lymphoprep density gradient medium and SepMate-50 tubes. Isolated T cells were cryopreserved at 2×10^7 T cells per vial in CryoStor CS10 cryopreservation medium (Stem Cell Technologies).

CAR T cell production

Cryopreserved T cells were thawed and activated same day with Human T-Expander CD3/CD28 Dynabeads (Gibco) at 3:1 beads:cell ratio in T cell medium (AIMV supplemented with 5% fetal bovine serum (FBS), 10 mM HEPES, 2 mM GlutaMAX, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin (Gibco)). Recombinant human IL-2 (Peprotech) was provided at 100 U ml⁻¹. T cells were transduced with retroviral vector on days 2 and 3 after activation and maintained at 0.5×10^6 – 1×10^6 cells per ml in T cell medium with IL-2. Unless otherwise indicated, CAR T cells were used for in vitro assays or transferred into mice on days 10–11 after activation.

Retroviral transduction

Non-tissue culture treated 12-well plates were coated overnight at 4 °C with 1 ml Retronectin (Takara) at 25 µg ml⁻¹ in PBS. Plates were washed with PBS and blocked with 2% BSA for 15 min. Thawed retroviral supernatant was added at approximately 1 ml per well and centrifuged for 2 h at 32 °C at 3,200 rpm before the addition of cells.

CRISPR knockout

CRISPR–Cas9 gene knockout was performed by transient Cas9/gRNA (RNP) complex electroporation using the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza). On day 4 of culture, HA-28z CAR T cells were counted, pelleted and resuspended in P3 buffer at 1.5×10^6 – 2×10^6 cells per 20 µl reaction. 3.3 µg Alt-R .Sp. Cas9 protein (IDT) and 40 pmol chemically modified synthetic sgRNA (Synthego) (2:1 molar ratio gRNA:Cas9) per reaction was pre-complexed for 10 min at room temperature to create ribonucleoprotein complexes (RNP). A 20-µl cell suspension was mixed with RNP and electroporated using the EO-115 protocol in 16-well cuvette strips. Cells were recovered at 37 °C for 30 min in 200 µl T cell medium then expanded as described above. Knockdown efficiency was determined using TIDE and/or immunoblot. Control HA-28z CAR T cells were electroporated with a gRNA targeting the safe-harbour locus AAVS1. The following gRNA target sequences were used: AAVS1- GGGGCCACTAGGGACAGGAT, JUNB-ACTCCTGAAACCGAGC CTGG, BATF-TCAGTCTGTCGGAGCTGTG, BATF3-CGTCCTGCAGAG GAGCGTCG, and IRF4-CGGAGAGTTCGGCATGAGCG.

Cell lines

The Kelly neuroblastoma, EW8 Ewing's sarcoma, 143b and TC32 osteosarcoma cell lines were originally obtained from ATCC. In some cases, cell lines were stably transduced with GFP and firefly luciferase (GL). The CD19⁺CD22⁺ Nalm6-GL B-ALL cell line was provided by D. Barrett. Nalm6-GD2 was created by co-transducing Nalm6-GL with cDNAs for GD2 synthase and GD3 synthase. The Nalm6-Her2 cell line was created using lentiviral overexpression of Her2 cDNA. Single-cell clones were then chosen for high antigen expression. Nalm6-22-knockout (Nalm6-22^{KO}) and Nalm6-22^{KO} plus CD22^{low} (N6-22^{low}) have been previously described and were provided by T. Fry¹³. The Nalm6-CD19^{low} cell lines were created by R. Majzner (manuscript in preparation). All cell lines were cultured in complete media (CM) (RPMI supplemented with 10% FBS, 10 mM HEPES, 2 mM GlutaMAX, 100 U ml⁻¹ penicillin, and 100 µg ml⁻¹ streptomycin (Gibco)). STR DNA profiling of all cell lines is conducted by Genetica Cell Line testing once per year. None of the cell lines used in this study is included in the commonly misidentified cell lines registry. Before using for in vivo experiments, cell lines were tested with MycoAlert detection kit (Lonza). All cell lines tested negative.

Flow cytometry

The anti-CD19 CAR idiotype antibody was provided by B. Jena and L. Cooper⁴⁸. The 1A7 anti-14G2a idiotype antibody was obtained from NCI-Frederick. CD22 and Her2 CARs were detected using human CD22-Fc and Her2-Fc recombinant proteins (R&D). The idiotype antibodies and Fc-fusion proteins were conjugated in house with Dylight488 and/or 650 antibody labelling kits (Thermo Fisher). T cell surface phenotype was assessed using the following antibodies:

From BioLegend: CD4-APC-Cy7 (clone OKT4), CD8-PerCp-Cy5.5 (clone SK1), TIM-3-BV510 (clone F38-2E2), CD39-FITC or APC-Cy7 (clone A1), CD95-PE (clone DX2), CD3-PacBlue (clone HIT3a).

From eBioscience: PD-1-PE-Cy7 (clone eBioJ105), LAG-3-PE (clone 3DS223H), CD45RO-PE-Cy7 (clone UCHL1), CD45-PerCp-Cy5.5 (clone HI30).

From BD: CD45RA-FITC or BV711 (clone HI100), CCR7-BV421 (clone 150503), CD122-BV510 (clone Mik-β3), CD62L-BV605 (clone DREG-56), CD4-BUV395 (clone SK3), CD8-BUV805 (clone SK1).

Cytokine production

Approximately 1×10^5 CAR⁺ T cells and 1×10^5 tumour cells were cultured in 200 μ l CM in 96-well flat bottom plates for 24 h. For idiotype stimulation, serial dilutions of 1A7 were crosslinked in 1 \times Coating Buffer (BioLegend) overnight at 4 °C on Nunc Maxisorp 96-well ELISA plates (Thermo Scientific). Wells were washed once with PBS and 1×10^5 CAR⁺ T cells were plated in 200 μ l CM and cultured for 24 h. Triplicate wells were plated for each condition. Culture supernatants were collected and analysed for IFN γ and IL-2 by ELISA (BioLegend).

Intracellular cytokine staining

For intracellular cytokine staining analysis, CAR⁺ T cells and target cells were plated at 1:1 effector:target ratio in CM containing 1 \times monensin (eBioscience) and 5 μ l per test CD107a antibody (BV605, Clone H4A3, BioLegend) for 5–6 h. After incubation, intracellular cytokine staining was performed using the FoxP3 TF Staining Buffer Set (eBioscience) according to the manufacturer's instruction using the following antibodies from BioLegend: IL2-PECy7 Clone MQ1-17H12, IFN γ -APC/Cy7 Clone 4S.B3, and TNF α -BV711 Clone Mab11.

Incucyte lysis assay

Approximately 5×10^4 GFP⁺ leukaemia cells were co-cultured with CAR T cells in 200 μ l CM in 96-well flat bottom plates for up to 120 h. Triplicate wells were plated for each condition. Plates were imaged every 2–3 h using the IncuCyte ZOOM Live-Cell analysis system (Essen Bioscience). Four images per well at 10 \times zoom were collected at each time point. Total integrated GFP intensity per well was assessed as a quantitative measure of live, GFP⁺ tumour cells. Values were normalized to the starting measurement and plotted over time. Effector:target ratios are indicated in the figure legends.

Immunoblotting and immunoprecipitations

Whole-cell protein lysates were obtained in non-denaturing buffer (150 mM NaCl, 50 mM Tris pH 8, 1% NP-10, 0.25% sodium deoxycholate). Protein concentrations were estimated by Bio-Rad colorimetric assay. Immunoblotting was performed by loading 20 μ g of protein onto 11% PAGE gels followed by transfer to PVF membranes. Signals were detected by enhanced chemiluminescence (Pierce) or with the Odyssey imaging system. Representative blots are shown. The following primary antibodies used were purchased from Cell Signaling: c-Jun (60A8), P-c-Jun^{Ser73} (D47G9), JunB (C37F9), BATF (D7C5), IRF4 (4964) and Histone-3 (1B1B2). The BATF3 (AF7437) antibody was from R&D. Immunoprecipitations were performed in 100 μ g of whole-cell protein lysates in 150 μ l of nondenaturing buffer and 7.5 μ g of agar-conjugated antibodies c-Jun (G4) or JunB (C11) (Santa Cruz Biotechnology). After overnight incubation at 4 °C, beads were washed three times with nondenaturing buffer, and proteins were eluted in Laemmli sample buffer, boiled and loaded onto PAGE gels. Detection of immunoprecipitated proteins was performed with above-mentioned reagents and antibodies.

Preparation of chromatin fractions

Separation of chromatin-bound from soluble proteins was performed as previously described⁴⁹ using cytoskeletal (CSK) buffer: 10 mM PIPES-KOH (pH 6.8), 100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂, 0.5 mM PMSF, 0.1 mM glycerolphosphate, 50 mM NaF, 1 mM Na₃VO₄, containing 0.1% Nonidet P-40 and protease inhibitors 2 mM PMSF, 10 μ g ml⁻¹ leupeptin, 4 μ g ml⁻¹ aprotinin, and 4 μ g ml⁻¹ pepstatin. In brief, cell pellets were lysed for 10 min on ice followed by 5,000 rpm centrifugation at 4 °C for 5 min. The soluble fraction was collected and cleared by high-speed centrifugation, 13,000 rpm for 5 min. Protein concentration was determined by Bradford assays. Pellets containing chromatin-bound proteins were washed with CSK buffer and centrifuged at 5,000 rpm at 4 °C for 5 min. Chromatin-bound proteins were solubilized in 1 \times Laemmli Sample Buffer and boiled for 5 min. Equal volumes of chromatin

and soluble fraction were loaded for each sample and analysed by immunoblotting.

ChIP and library preparation

Twenty-million CAR T cells were fixed with 1% formaldehyde for 10 min at room temperature. Cross-linking was quenched using 0.125 M glycine for 10 min before cells were washed twice with PBS. Cross-linked pellets were frozen with dry-ice ethanol and stored at -80 °C. Two biological replicates were collected for each cell culture. Chromatin immunoprecipitations were performed with exogenous spike-ins (ChIP-Rx) to allow for proper normalization, as previously described⁵⁰. In brief, pellets were thawed on ice before cell membrane lysis in 5 ml LB1 by rotating for 10 min at 4 °C. Nuclei were pelleted at 1,350g for 5 min at 4 °C and lysed in 5 ml LB2 by rotating for 10 min at room temperature. Chromatin was pelleted at 1,350g for 5 min at 4 °C and resuspended in 1.5 ml LB3. Sonication was performed in a Bioruptor Plus until chromatin was 200–700 bp. Debris were pelleted and supernatants were collected and Triton X-100 was added to 1% final concentration. Ten per cent of the sample was collected as input controls. Anti-RF4 (Abcam Ab101168) or anti-c-Jun (Active Motif 39309) targeting antibodies were added at 5 μ g per immunoprecipitate to sonicated lysate and rotated at 4 °C for 16–20 h.

Protein G Dynabeads (100 μ l per immunoprecipitate) were washed three times with Block Solution (0.5% BSA in PBS). Antibody-bound chromatin was added to beads and rotated for 2–4 h at 4 °C. Bead-bound chromatin was washed five times with 1 ml RIPA wash buffer then once with 1 ml TE buffer with 500 mM NaCl. Beads were resuspended in 210 μ l. Elution and chromatin was eluted at 65 °C for 15 min. Beads were magnetized and supernatant was removed to a fresh tube. Immunoprecipitated and input control chromatin was reverse cross-linked at 65 °C for 12–16 h.

Samples were diluted with 1 volume TE buffer. RNA was digested using 0.2 mg ml⁻¹ RNase A (Qiagen 19101) for 2 h at 37 °C. CaCl₂ was added to 5.25 mM and samples were treated with 0.2 mg ml⁻¹ proteinase K (Life Technologies E00491) for 30 min at 55 °C. One volume phenol-chloroform-isoamyl alcohol was added and centrifuged 16,500g for 5 min to extract DNA, followed by a second extraction using one volume pure chloroform. Aqueous phase was removed and DNA was precipitated using two volumes ethanol and 0.3 M sodium acetate. DNA pellets were resuspended in EB elution buffer (Qiagen).

To prepare libraries for sequencing, DNA was end repaired using T4 polymerase (New England Biolabs M0203L), Klenow fragment (NEB M0210L), and T4 polynucleotide kinase (NEB M0201L) for 30 min at 20 °C. 3' A-tailing was performed using Exo- Klenow fragment (NEB M0212L) for 30 min at 37 °C. Illumina TruSeq Pre-Indexed Adaptors (1 μ M) or NEBNext Illumina Multiplex Oligo Adaptors (NEB E7335S) were ligated for 1 h at room temperature. Unligated adapters were separated by gel electrophoresis (2.5% agarose, 0.5 \times TBE) and ligated DNA was purified using a NucleoSpin Gel Clean-up Kit (Macherey-Nagel 740609.250). Ligated DNA was PCR amplified using TruSeq Primers 1.0 and 2.0 or NEBNext Multiplex Primers and purified using AMPure XP beads (Beckman Coulter A63881). Purified libraries were quantified using Agilent 2100 Bioanalyzer HS DNA and multiplexed in equimolar concentrations. Sequencing was performed using an Illumina NextSeq or HiSeq at 2 \times 75 bp by Stanford Functional Genomics Facility.

Mice

Immunocompromised NOD-SCID-*Il2rg*^{-/-} (NSG) mice were purchased from JAX and bred in-house. All mice were bred, housed and treated in ethical compliance with Stanford University IACUC (APLAC) approved protocols. Six-to-eight-week-old male or female mice were inoculated with either 1×10^6 Nalm6-GL leukaemia via intravenous or 0.5×10^6 – 1×10^6 143B osteosarcoma via intramuscular injections. All CAR T cells were injected intravenously. Time and treatment dose are indicated in the figure legends. Leukaemia progression was measured

by bioluminescent imaging using the IVIS imaging system. Values were analysed using Living Image software. Solid tumour progression was followed using caliper measurements of the injected leg area. Mice were humanely euthanized when an IACUC-approved end-point measurement reached 1.75 cm in either direction (for solid tumour) or when mice demonstrated signs of morbidity and/or hind-limb paralysis (leukaemia). Five-to-ten mice per group were treated in each experiment based on previous experience in these models, and each experiment was repeated two or three times as indicated. Mice were randomized to ensure equal pre-treatment tumour burden before CAR T cell treatment. In some experiments, researchers were blinded to treatment during tumour measurement.

Blood and tissue analysis

Peripheral blood sampling was conducted via retro-orbital blood collection under isoflurane anaesthesia at the indicated time points. Fifty microlitres of blood was labelled with CD45, CD3, CD4 and CD8, lysed using BD FACS Lysing Solution and quantified using CountBright Absolute Counting beads (Thermo Fisher) on a BD Fortessa flow cytometer. For ex vivo analysis of CAR TILs, 14 days after T cell treatment (day 28 after tumour engraftment), six mice per group were euthanized, solid tumour tissue was collected, mechanically dissociated using the gentleMACS dissociator (Miltenyi), and single-cell suspensions were either analysed by flow cytometry, re-plated with 3×10^5 Nalm6-Her2⁺ target cells for ICS analysis, or labelled for sorting. Live, CD45⁺ TILs were sorted from each tumour and re-stimulated at 1:1 effector:target ratio with Nalm6-Her2⁺ target cells. Twenty-four-hour supernatant was analysed for IL-2 production by ELISA.

ATAC-seq

ATAC-seq library preparation was carried out as previously described⁵¹. In brief, 100,000 cells from each sample were sorted by FACS into CM, centrifuged at 500g at 4 °C, then resuspended in ATAC-seq resuspension buffer (RSB) (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl₂) supplemented with 0.1% NP-40, 0.1% Tween-20 and 0.01% digitonin. Samples were split into two replicates each before all subsequent steps. Samples were incubated on ice for 3 min, then washed out with 1 ml RSB supplemented with 0.1% Tween-20. Nuclei were pelleted at 500g for 10 min at 4 °C. The nuclei pellet was resuspended in 50 µl transposition mix (25 µl 2× TD buffer, 2.5 µl transposase (Illumina), 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20, 5 µl H₂O) and incubated at 37 °C for 30 min in a thermomixer with 1,000 rpm shaking. The reaction was cleaned up using the Qiagen MinElute PCR Purification Kit. Libraries were PCR-amplified using the NEBNext Hi-Fidelity PCR Master Mix and custom primers (IDT) as previously described²³. Libraries were sufficiently amplified following 5 cycles of PCR, as indicated by qPCR fluorescence curves²³. Libraries were purified with the Qiagen MinElute PCR Purification Kit and quantified with the KAPA Library Quantification Kit. Libraries were sequenced on the Illumina NextSeq at the Stanford Functional Genomics Facility with paired-end 75-bp reads. Adaptor sequences were trimmed using SeqPurge and aligned to hg19 genome using bowtie2. These reads were then filtered for mitochondrial reads, low mapping quality ($Q \geq 20$), and PCR duplicates using Picard tools. Then we converted the bam to a bed and got the Tn5 corrected insertion sites ('+' stranded + 4 bp, '-' stranded -5 bp). To identify peaks, we called peaks for each sample using MACS2⁻shift -75-extsize 150-nomodel-call-summits-nolambda-keep-dup all -p 0.00001¹ using the insertion beds. To get a union peak set, we (1) extended all summits to 500 bp; (2) merged all summit bed files; and (3) used bedtools cluster and selected the summit with the highest MACS2 score. This was then filtered by the ENCODE hg19 blacklist (<https://www.encodeproject.org/annotations/ENCSR636HFF/>), to remove peaks that extend beyond the ends of chromosomes. We then annotated these peaks using HOMER and computed the occurrence of a transcription factor motif using motifmatchr in R with chromVARMotifs HOMER set. To create sequencing

tracks, we read the Tn5 corrected insertion sites into R and created a coverage pileup binned every 100 bp using rtracklayer. We then counted all insertions that fell within each peak to get a counts matrix (peak × samples). To determined differential peaks we first used peaks that were annotated as 'TSS' as control genes or 'housekeeping peaks' for DESeq2 and then computed differential peaks with this normalization. All clustering was performed using the regularized log transform values from DESeq2. Transcription factor motif deviation analysis was carried out using chromVAR as previously described²⁴. Transcription factor motif enrichment were calculated using a hypergeometric test in R testing the representation of a motif (from motifmatchr above) in a subset of peaks vs all peaks.

Subset RNA-seq

For T cell subset-specific RNA-seq, T cells were isolated from healthy donor buffy coats as described above. Before activation, naive and central memory CD4⁺ or CD8⁺ subsets were isolated using a BD FACS Aria cell sorter (Stem Cell FACS Core, Stanford University School of Medicine) using the following markers: naive (CD45RA⁺CD45RO⁻, CD62L⁺, CCR7⁺, CD95⁻, and CD122⁻), central memory (CD45RA⁻CD45RO⁺, CD62L⁺, CCR7⁺). Sorted starting populations were activated, transduced and cultured as described above. On days 7, 10 and 14 of culture, CAR⁺ CD4⁺ and CD8⁺ cells were sorted, and RNA was isolated using Qiagen mRNEasy kit. Samples were library prepped and sequenced via Illumina NextSeq paired end platform by the Stanford Functional Genomics Core.

Bulk RNA-seq

For bulk RNA isolation, healthy donor T cells were prepared as described. On day 10 or 11 of culture, total mRNA was isolated from 2×10^6 bulk CAR T cells using Qiagen RNEasy Plus mini isolation kit. Bulk RNA-seq was performed by BGI America (Cambridge, MA) using the BGISEQ-500 platform, single-end 50-bp read length, at 30×10^6 reads per sample. Principal component analysis was performed using stats package and plots with ggplot2 package in R (version 3.5)⁵². GSEA was performed using the GSEA software (Broad Institute) as described^{53,54}. DAVID analysis was performed for transcription factor enrichment as described^{55,56}.

Single-cell RNA-seq

To compare gene expression in single CD19-28z and GD2-28z CAR cells, we sorted naive T-cell subset on day 0 for subsequent single-cell analysis on day 10 using the Chromium platform (10X Genomics) and the Chromium Single Cell 3' v2 Reagent Kit according to the manufacturer's instructions. cDNA libraries were prepared separately for CD19-CAR and GD2-CAR cells, and the CD4⁺ cells and CD8⁺ cells were combined in each run to be separated bioinformatically downstream. Sequencing was performed on the Illumina NextSeq system (paired-end, 26 bp into read 1 and 98 bp into read 2) to a depth of more than 100,000 reads per cell. Single-cell RNA-seq reads were aligned to the Genome Reference Consortium Human Build 38 (GRCh38), normalized for batch effects, and filtered for cell events using the Cell Ranger software (10X Genomics). A total of 804 CD19-CAR and 726 GD2-CAR T cells were sequenced to an average of 350,587 post-normalization reads per cell. The cell-gene matrix was further processed using the Cell Ranger RKit software (10X Genomics) as previously described⁵⁷. In brief, we first selected genes with at least one unique molecular identifier (UMI) counts in any given cell. UMI counts were then normalized to UMI sums for each cell and multiplied by a median UMI count across cells. Next, the data were transformed by taking a natural logarithm of the resulting data matrix. For correlation network of exhaustion-related transcription factors, transcription factor genes identified as differentially expressed ($P < 0.05$) by DESeq2 form the nodes of the network. Colours represent log₂-transformed fold change (GD2 vs CD19 CAR). Edge thickness represents the magnitude of correlation in

Article

expression between the relevant pair of genes across cells. Correlation score greater than 0.1 was used to construct networks.

To compare gene expression in single JUN-overexpressing and control Her2-BBz CAR T cells in vivo, live human CD45⁺ tumour-infiltrating cells were sorted and pooled from six NSG mice bearing 143B osteosarcoma tumours 14 days after CAR T cell infusion. Sorted cells were analysed using the 10X Genomics platform as described above and sequenced on the Illumina HighSeq 4000 system to a depth of more than 50,000 reads per cell. A total of 6,946 Her2-BBz and 10,985 JUN-Her2-BBz cells were sequenced to an average of 49,542 post-normalization reads per cell. The cell–gene matrix was further processed using the Seurat v.3.0 software^{58,59}. In brief, we selected genes expressed in ≥50 cells. Single live cells were selected as droplets expressing ≥500 genes with ≤20,000 UMI counts and ≤10% mitochondrial reads. UMI count data matrix was transformed and scaled, including variable feature selection, with SCTransform pipeline. T cells were selected as CD3⁺ events (99.3% cells expressing *CD3G*, *CD3D*, *CD3E*, and/or *CD247* gene). Where indicated, CD4⁺ and CD8⁺ T cell subsets were selected (8.3% CD4⁺ CD8[−], 70.3% CD4[−] CD8⁺). The resulting data matrix was then examined using differential expression analysis, cell cycle analysis, clustering, and UMAP embedding.

Statistical analysis

Unless otherwise noted, statistical analyses for significant differences between groups were conducted using unpaired two-tailed *t*-tests without correction for multiple comparisons and without assuming consistent s.d. using GraphPad Prism 7. Survival curves were compared using the log-rank Mantel–Cox test. See Supplementary Table 2 for full statistical analyses, including exact *P* values, *t*-ratio, and degrees of freedom.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The sequencing datasets generated in this publication have been deposited in NCBI Gene Expression Omnibus (GEO)^{60,61} and are accessible through GEO series accession numbers: bulk RNA-seq: GSE136891, scRNA-seq CD19/GD2-28z: GSE136874, scRNA-seq control/JUN-Her2-BBz TILs: GSE136805, ATAC-seq: GSE136796, ChIP-seq: GSE136853.

47. Hudecek, M. et al. The nonsignaling extracellular spacer domain of chimeric antigen receptors is decisive for in vivo antitumor activity. *Cancer Immunol. Res.* **3**, 125–135 (2015).
48. Jena, B. et al. Chimeric antigen receptor (CAR)-specific monoclonal antibody to detect CD19-specific T cells in clinical trials. *PLoS ONE* **8**, e57838 (2013).

49. Sotillo, E. et al. Coordinated activation of the origin licensing factor CDC6 and CDK2 in resting human fibroblasts expressing SV40 small T antigen and cyclin E. *J. Biol. Chem.* **284**, 14126–14135 (2009).
50. Nagaraja, S. et al. Transcriptional dependencies in diffuse intrinsic pontine glioma. *Cancer Cell* **31**, 635–652 (2017).
51. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
52. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
53. Mootha, V. K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
54. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
55. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
56. Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
57. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
58. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
59. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
60. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
61. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).

Acknowledgements This work was supported by a Stand Up To Cancer–St Baldrick’s–National Cancer Institute Pediatric Dream Team Translational Cancer Research Grant (C.L.M.), the Parker Institute for Cancer Immunotherapy (C.L.M., H.Y.C., Z.G.), the Virginia and D.K. Ludwig Fund for Cancer Research (C.L.M.), and NIH P50-HG007735 (H.Y.C.). H.Y.C. is an Investigator of the Howard Hughes Medical Institute. A.T.S. was supported by a Parker Bridge Scholar Award from the Parker Institute for Cancer Immunotherapy and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. R.C.L. was supported by the Emerson Collective Cancer Research Fund. The Illumina HiSeq 4000 used here was purchased with the NIH funds (award S10OD018220). Figure 4a was created by S. Knemeyer, SciStories LLC.

Author contributions R.C.L. cloned the constructs, designed and performed experiments, analysed data, and wrote the manuscript. E.W.W. and E.S. designed and performed experiments. E.S. performed all immunoblots, immunoprecipitations and GSEAs. D.G., J.G., A.T.S. and H.Y.C. performed and analysed ATAC-seq. Z.G., C.F.A.d.B. and S.R.Q. performed and analysed single-cell RNA-seq. H.A. and R.J. performed and analysed bulk RNA-seq. J.L. and V.T. cloned the JUN-mutant and JUN-DD constructs and performed experiments. R.M. cloned the HA-GD2 CAR and created the CD19^{low} Nalm6. P.X. performed mouse injections and imaging. S.N. and M.M. performed ChIP-seq experiments and analysis. C.L.M. designed experiments and wrote the manuscript.

Competing interests C.L.M., R.C.L., E.W.W. and E.S. are inventors on a Stanford University Provisional patent pending on modulating AP-1 to enhance function of T cells; 62/599,299; C.L.M. is a founder of, holds equity in and receives consulting fees from Lyell Immunopharma, which has licensed the technology. R.C.L. is employed by and E.W.W. and E.S. are consultants for Lyell Immunopharma.

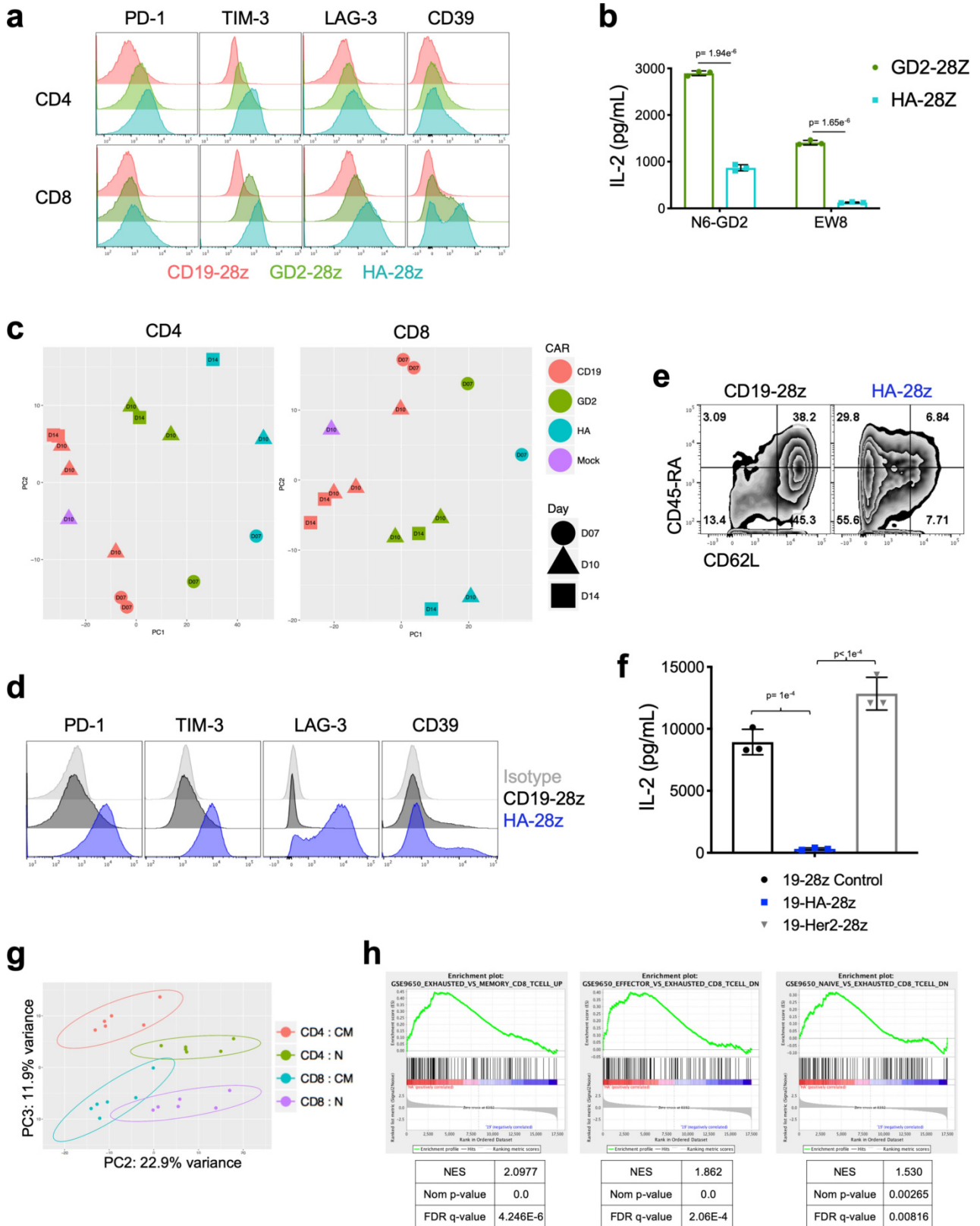
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1805-z>.

Correspondence and requests for materials should be addressed to C.L.M.

Peer review information Nature thanks Steven Albelda, Takeshi Egawa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

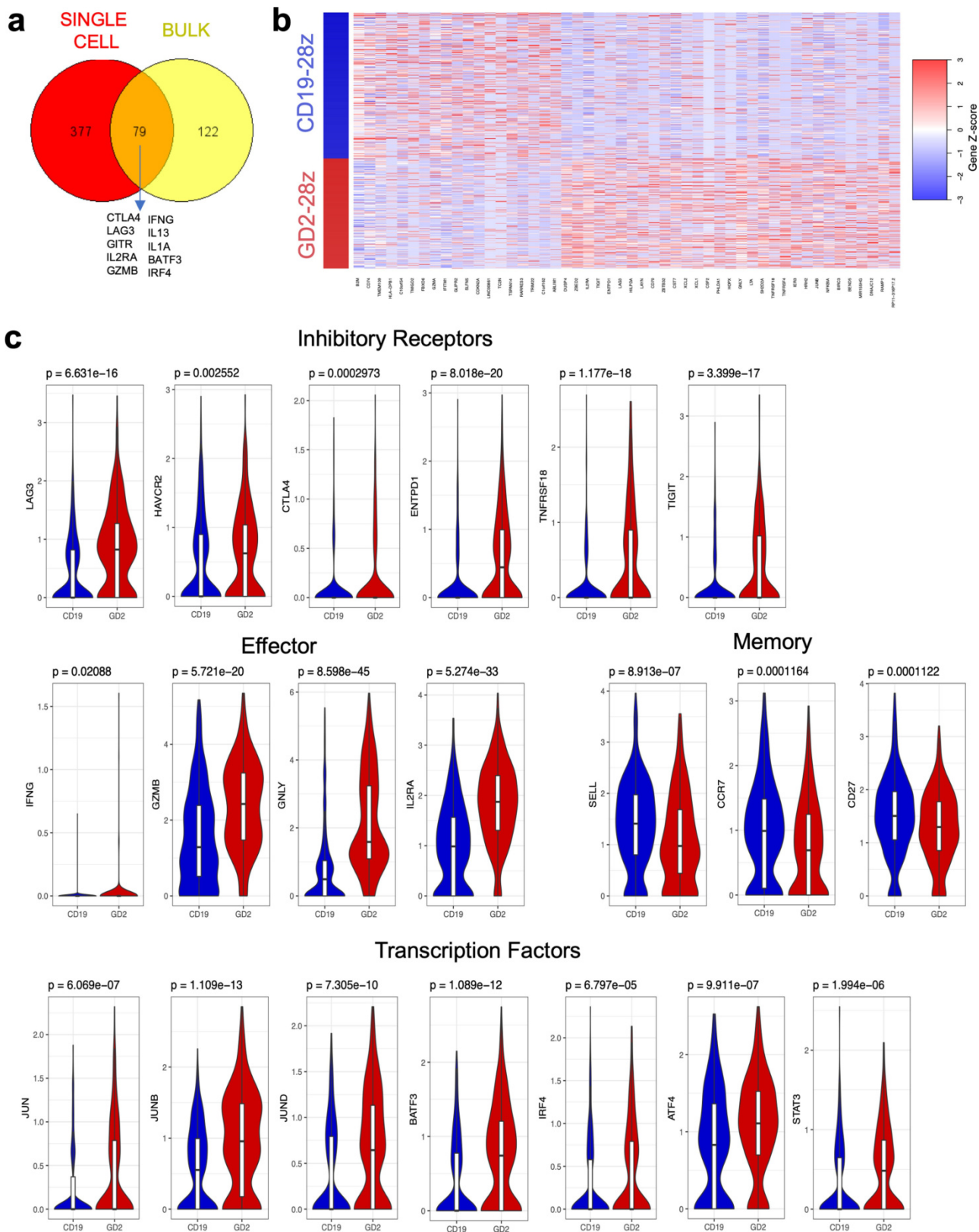


Extended Data Fig. 1 | See next page for caption.

Article

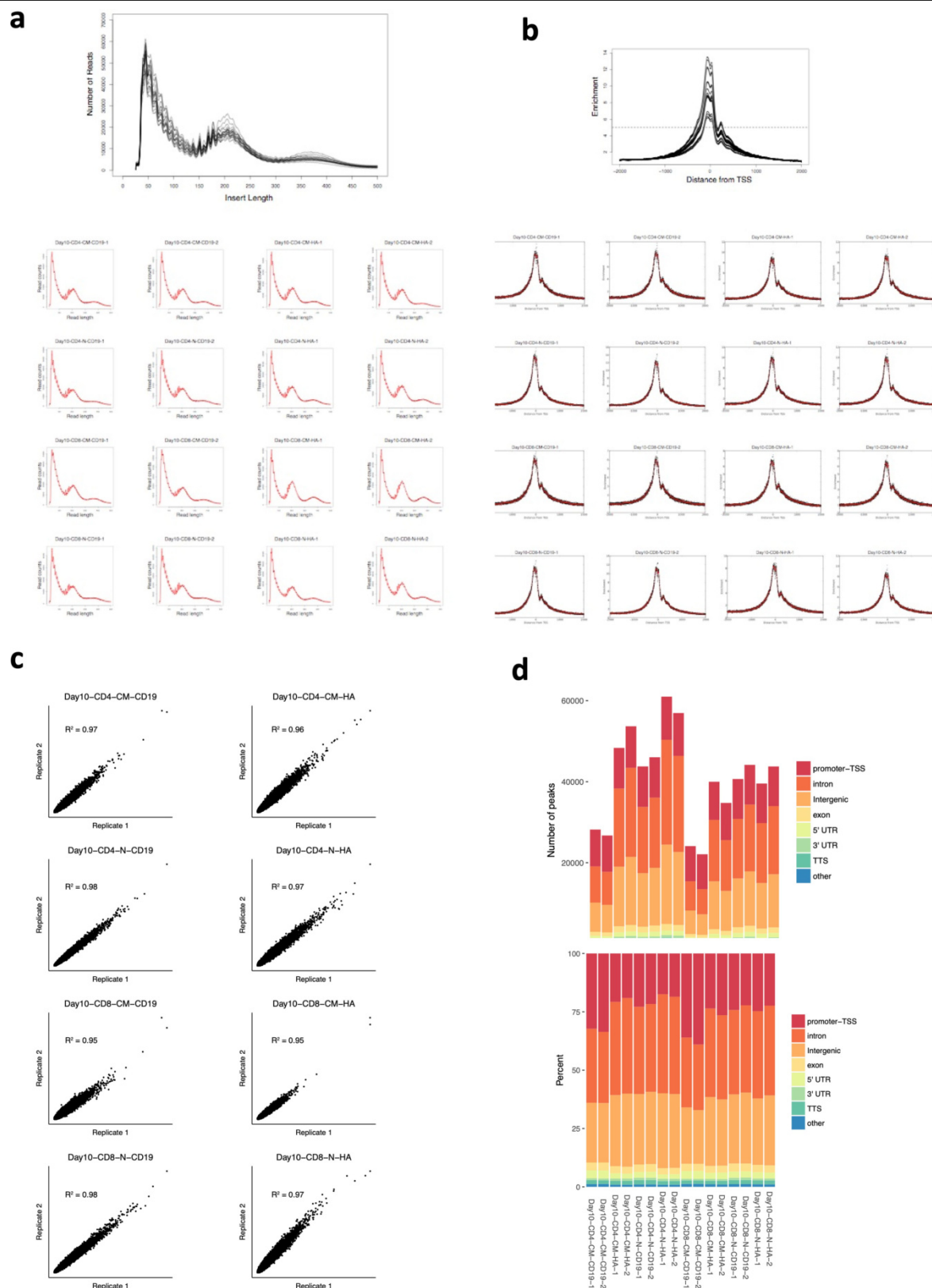
Extended Data Fig. 1 | High-affinity 14g2a-GD2(E101K) CAR T cells manifest an exaggerated exhaustion signature compared with the original 14g2a-GD2 CAR T cells. **a**, Surface inhibitory receptor expression in CD19, GD2 and HA-GD2(E101K) CAR T cells at day 10 of culture. High-affinity E101K mutation results in increased inhibitory receptor expression in CD4⁺ and CD8⁺ CAR T cells, compared with parental GD2 CAR T cells. **b**, IL-2 secretion after 24h co-culture of HA-GD2(E101K) or original GD2-28z CAR T cells with GD2⁺ target cells. The increased exhaustion profile of HA-GD2(E101K) CAR T cells corresponds to decreased functional activity, as measured by the ability to produce IL-2 after stimulation. Data are mean \pm s.d. of triplicate wells; representative of four independent experiments. **c**, PCA of bulk RNA-seq demonstrates larger variance between HA-GD2(E101K) and CD19 CAR T cells, whereas GD2-28z (short hinge) CAR T cells are intermediary. Left, CD4⁺ T cells. Right, CD8⁺ CAR T cells, naive-derived. Number of replicates is indicated. **d, e**, HA-GD2(E101K) CAR expression causes enhanced inhibitory receptor expression (**d**) and decreased memory formation (**e**) in CD4⁺ CAR T cells.

(See Fig. 1 for CD8⁺ data). **f**, IL-2 secretion from control CD19-28z CAR T cells or CD19 CAR T cells with bi-cistronic expression of tonically signalling HA-GD2(E101K) (19-HA-28z, blue) or bi-cistronic expression of Her2-28z (19-Her2-28z, grey) after 24-h stimulation with Nalm6 (CD19⁺GD2⁻Her2⁻) target cells to demonstrate that co-expression of HA-GD2-28z CAR induces T cell dysfunction in CD19-28z CAR T cells. Data are mean \pm s.d. of triplicate wells; representative of three independent experiments. **g**, RNA-seq PCA from Fig. 1 showing PC2 separation is driven by central memory versus naive starting subset and PC3 separation driven by CD4 versus CD8. **h**, Gene set enrichment analysis (GSEA): gene sets upregulated in day-10 HA-28z CAR T cells versus CD19-28z CAR T cells showed significant overlap with genes upregulated in exhausted versus memory CD8⁺ cells (left), exhausted versus effector CD8⁺ cells (middle), and exhausted versus naive CD8⁺ cells (right) in a mouse model of chronic viral infection⁴. NES, normalized enrichment score. In **b** and **f**, *P* values determined by unpaired two-tailed *t*-tests.



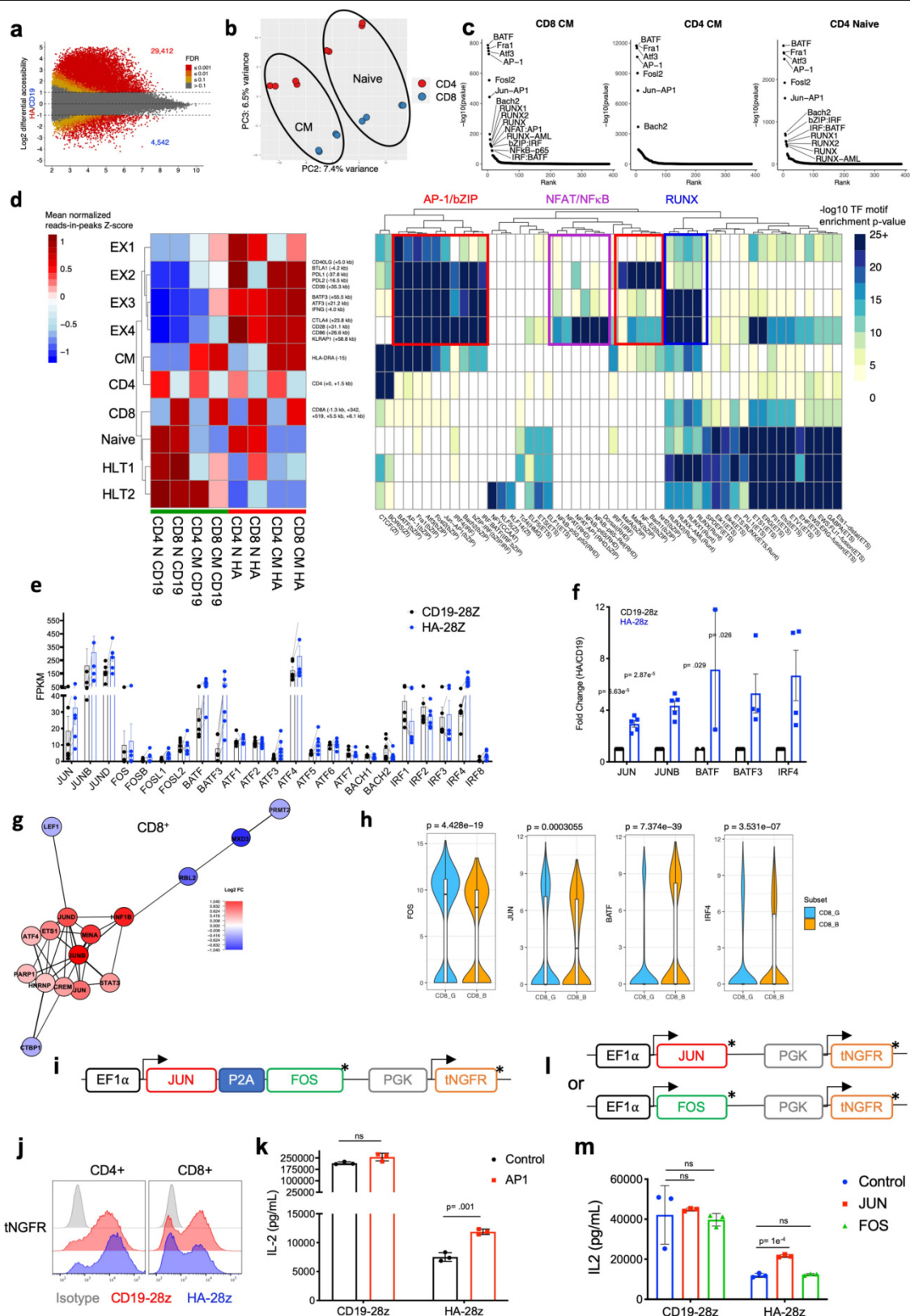
Extended Data Fig. 2 | GD2-28z CAR T cells display an exhaustion signature at the single cell-level. a, Venn diagram showing overlapping genes in differential expression analysis of single cell data (red) and the top 200 genes driving the separation of CD19 and HA-28z CAR T cells in bulk RNA-seq (yellow, Fig. 1f). In total, 79 out of the top 200 genes from bulk RNA-seq are differentially expressed by DESeq2 analysis in GD2-28z versus CD19-28z single cells. Highlighted genes from the intersection include inhibitory receptors (*CTLA4*, *LAG3*, *GITR*), effector molecules (*CD25*, *IFNG*, *GZMB*), cytokines (*IL13* and *IL1A*)

and bZIP/IRF family transcription factors (*BATF3* and *IRF4*). **b**, Heat map clustering the top 50 differentially expressed genes in GD2-28z versus CD19-28z single-cell transcriptome analysis. Each row represents one cell. **c**, Violin plots depicting individual gene expression in CD8⁺ GD2-28z and CD19-28z single CAR T cells. Genes upregulated in GD2 CAR T cells include inhibitory receptors, effector molecules and AP-1 family transcription factors, whereas CD19 CAR T cells have increased expression of memory-associated genes. *P* values determined by unpaired two-tailed Wilcoxon–Mann–Whitney *U* test.



Extended Data Fig. 3 | ATAC-seq data quality control. a, b, Insert length (a) and insert distance (b) from transcriptional start site (TSS) for combined (top) and individual (bottom) samples. **c,** Correlation between replicate samples.

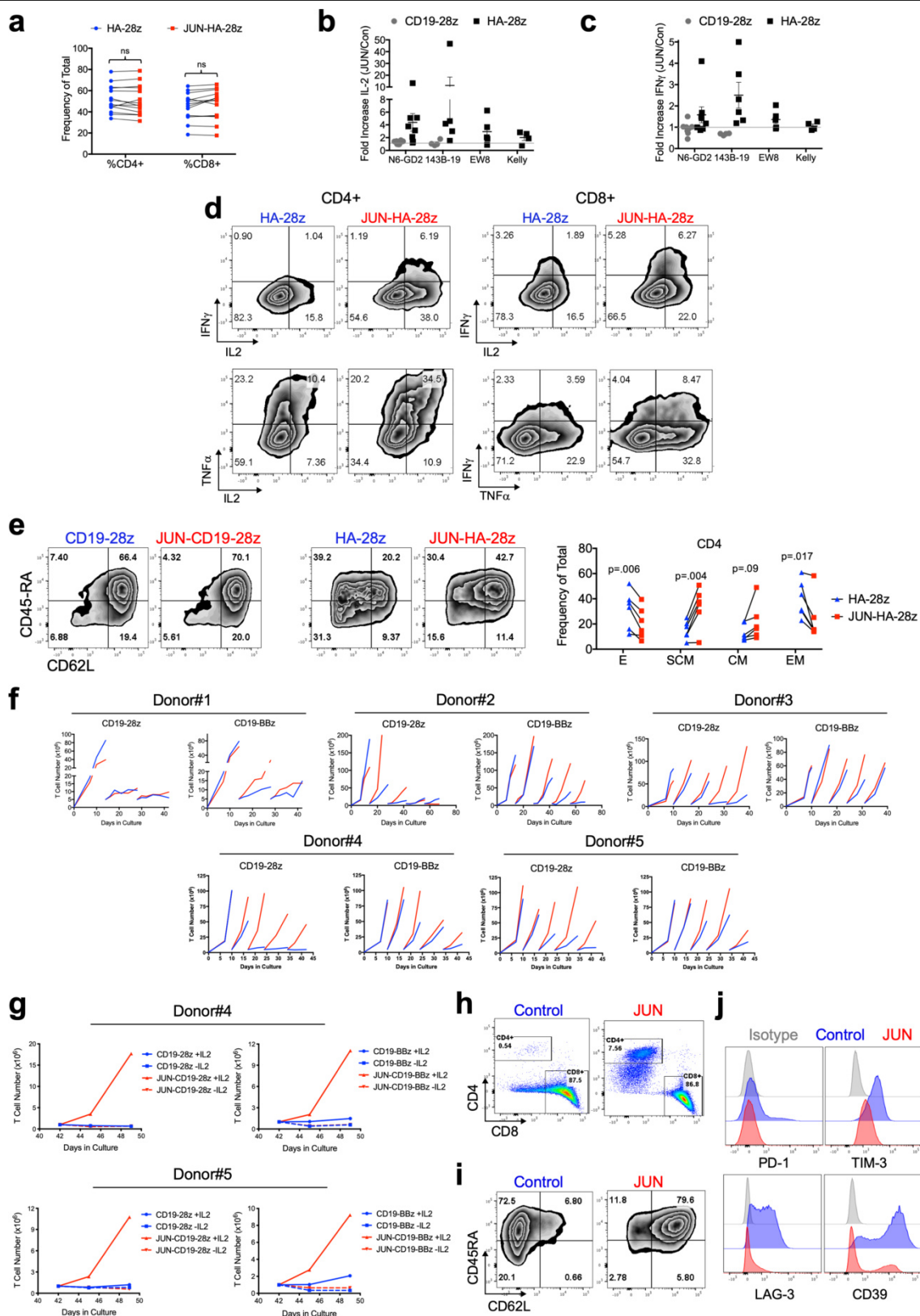
d, Location of mapped peaks in each sample by total number of peaks (top) and frequency of total (bottom).



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | AP-1 family transcription factors in HA-28z exhausted CART cells. **a**, Differentially accessible chromatin regions in CD4⁺ CD19-28z and HA-28z CAR T cells. Both the naive- and central-memory-derived subsets are incorporated for each CAR. **b**, PCA from Fig. 1h showing PC2 separation is driven by central memory versus naive, and PC3 separation driven by CD4 versus CD8. **c**, Top transcription factor motifs enriched in chromatin regions differentially accessible in HA-28z CAR T cells comprise AP-1–bZIP family factors in all starting T cell subsets. CD8⁺ naive subset is shown in Fig. 2. **d**, Peak clustering by shared regulatory motif (left) and enrichment heat map of transcription factor motifs (right) in each cluster. Ten different clusters including clusters associated with exhausted (EX1–EX4) or healthy (HLT1–HLT2) CAR T cells, central memory or naive starting subset, and CD4 or CD8 T cell subset. Genes of interest in each cluster are highlighted to the right. **e**, Bulk RNA-seq expression (fragments per kilobase of exon model per million reads mapped, FPKM) of indicated AP-1–bZIP and IRF family members in CD19-28z (black) and HA-28z (blue) CAR T cells. Data are mean \pm s.e.m. from $n = 6$ samples across three donors showing paired CD19 versus HA expression for each gene. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, Wilcoxon matched-pairs signed rank test (see Supplementary Information for exact P values). **f**, Increased protein expression of c-Jun, JunB, BATF3 and IRF4 in HA-28z versus CD19-28z CAR T cells at day 10 of culture by immunoblotting. Densitometry measurements of the fold change in HA compared with CD19 ($n = 2–5$ experiments). P values determined by unpaired two-tailed t -tests. **g**, Correlation network of exhaustion-related transcription factor in

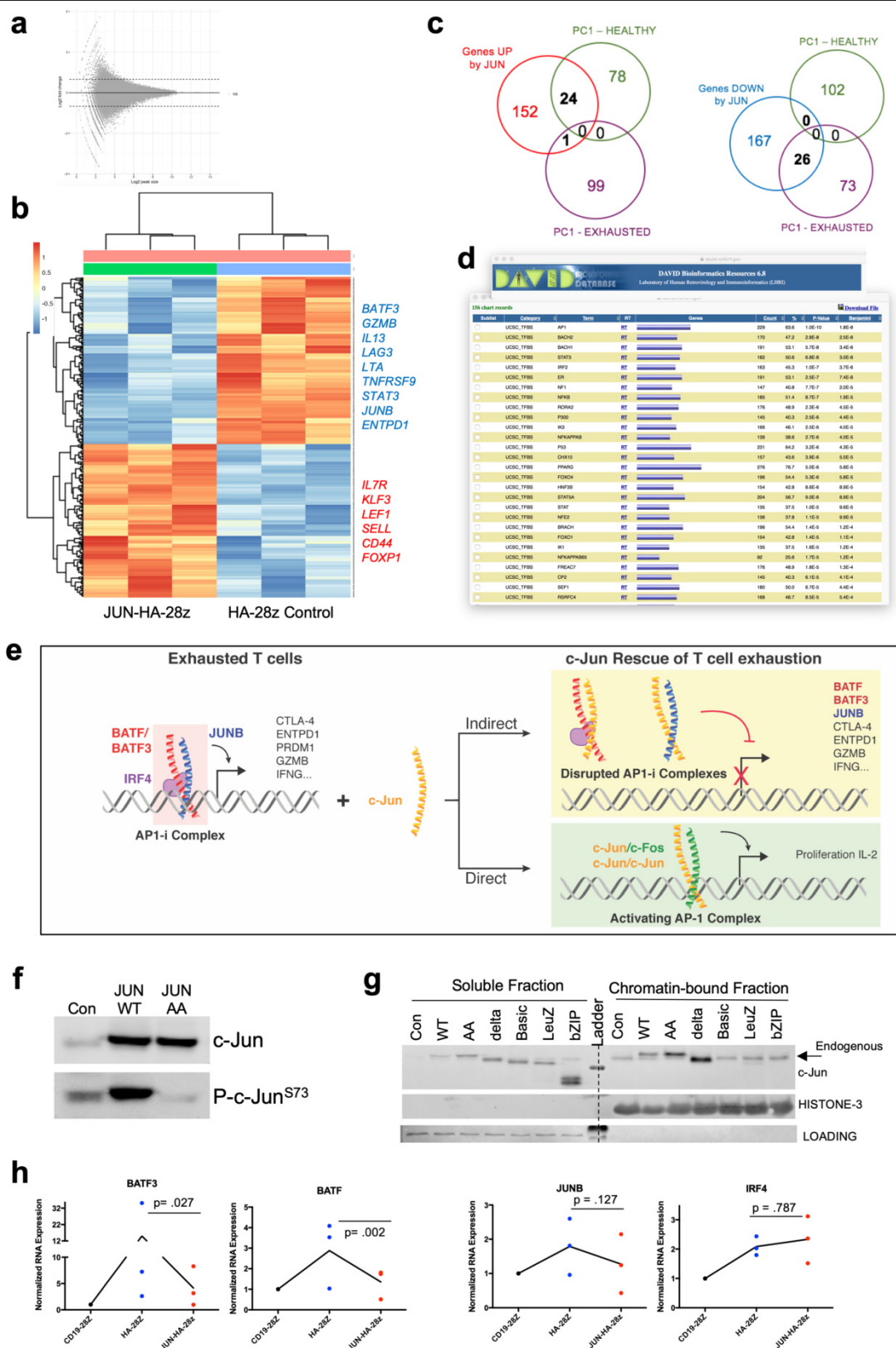
naive-derived CD8⁺ GD2-28z CAR T cells using single-cell RNA-seq analysis. **h**, Violin plots depicting single-cell gene expression of *FOS*, *JUN*, *BATF* and *IRF4* in CD8⁺ clusters associated with response (CD8.G) and non-response (CD8.B) in patients with metastatic melanoma after checkpoint therapy (CD8T-post-CD8G.B)²⁹. P values determined by unpaired two-tailed Wilcoxon–Mann–Whitney U test. In **i–m**, AP-1-modified HA-28z CAR T cells exhibit enhanced functional activity. **i–k**, CAR T cells were co-transduced with (AP-1) or without (control) a lentiviral vector encoding AP-1 transcription factors Fos and c-Jun and a truncated NGFR (tNGFR) surface selection marker. **i**, Schematic of the lentiviral construct. **j**, Representative transduction efficiency of AP-1-modified CAR T cells as measured by NGFR surface expression in indicated CD4⁺ and CD8⁺ CAR T cells. **k**, IL-2 production in control (black) or AP-1-modified (red) CAR T cells after 24-h stimulation with 143B-CD19 target cells. AP-1-modified HA-28z CAR T cells show increased IL-2 production compared with control CAR T cells. Data are mean \pm s.d. of triplicate wells; representative of two independent experiments. **l, m**, CAR T cells were co-transduced with lentiviral vectors encoding either c-Fos or c-Jun and a tNGFR surface selection marker. **l**, Schematics of the c-Fos and c-Jun lentiviral constructs. **m**, IL-2 production in control (blue), Fos (green) or c-Jun (red) modified CAR T cells after 24-h stimulation with Nalm6-GD2 target cells. Data are mean \pm s.d. of triplicate wells; representative of two independent experiments. Asterisk in **i** and **l** denotes a stop codon. P values determined by unpaired two-tailed t -tests. ns, not significant ($P > 0.05$).



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Enhanced activity of JUN-modified CAR T cells. JUN CAR T cells were produced as in Fig. 3. **a**, c-Jun overexpression does not affect the CD4:CD8 ratio in HA-28z CAR T cells at day 10 of culture ($n = 14$ independent experiments). Lines indicate paired samples from the same donor. P values determined by paired, two-tailed t -tests. **b, c**, Fold increase in IL-2 (**b**) and IFN γ (**c**) release after 24-h co-culture with the indicated target cells in JUN versus control CD19 and HA-28z CAR T cells. Each dot represents one independent experiment from different donors of $n = 8$ total experiments. **d**, Representative contour plots demonstrating increased intracellular cytokine production in both CD4 $^{+}$ and CD8 $^{+}$ JUN-HA-28z versus control HA-28z CAR T cells stimulated for 5 h with Nalm6-GD2 target cells. Representative of three independent experiments. **e**, Left, flow cytometry showing representative CD45RA and CD62L expression in control versus JUN CAR CD4 $^{+}$ T cells at day 10. Right, relative frequency of effector (CD45RA $^{+}$ CD62L $^{-}$), stem-cell memory (CD45RA $^{+}$ CD62L $^{+}$), central memory (CD45RA $^{-}$ CD62L $^{+}$), and effector memory (CD45RA $^{-}$ CD62L $^{-}$) cells in CD4 $^{+}$ control or JUN-HA-28z CAR T cells ($n = 6$ donors

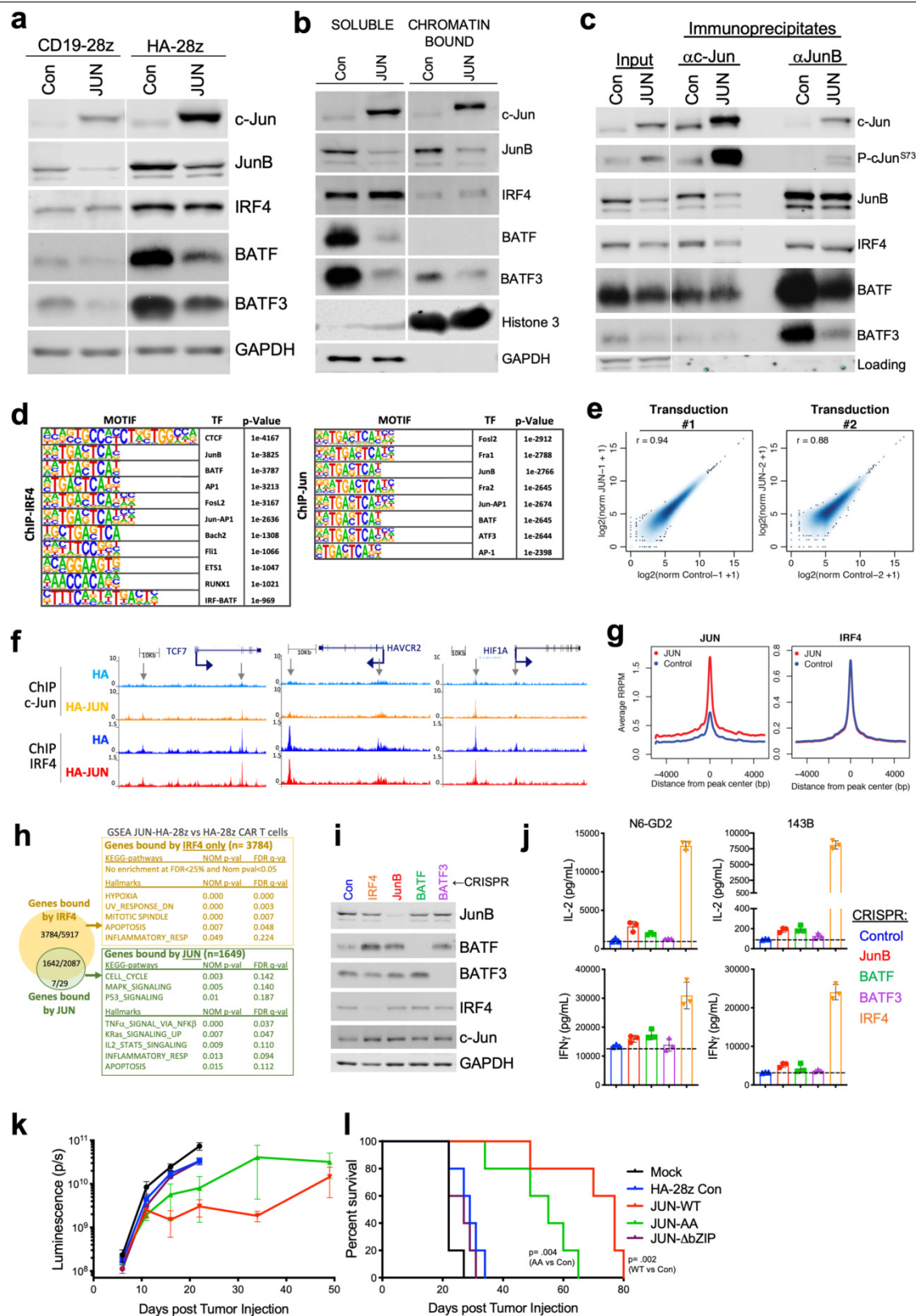
from independent experiments). Lines indicate paired samples from the same donor. P values determined by paired two-tailed t -tests. **f**, Extended in vitro expansion of control (blue) or JUN-modified (red) CD19 CAR T cells in five independent experiments with five different healthy donors. At the indicated time points, T cells were re-plated in fresh T cell media with 100 IU ml $^{-1}$ IL-2. T cells were counted and fed to keep cells at 0.5×10^6 cells per ml every 2–3 days. For Donor-1, 5×10^6 viable T cells were re-plated on days 14 and 28. For Donor-2, 5×10^6 viable T cells were re-plated on days 14, 28, 42 and 56. For Donor-3, 5×10^6 viable T cells were re-plated on days 10, 17, 24 and 31. For Donor-4 and Donor-5, 5×10^6 viable T cells were re-plated on days 10, 17, 24 and 34. **g**, On day 42 of culture, 1×10^6 viable T cells from Donor-4 (top) and Donor-5 (bottom) were re-plated and cultured for 7 days with or without IL-2. **h–j**, Cell-surface phenotype of control or JUN-CD19-28z CAR T cells from Fig. 3g (Donor-3) on day 46 of culture. **h**, CD4 versus CD8 expression. **i**, Surface expression of CD45RA versus CD62L. **j**, Day 46 surface exhaustion marker expression in CD8 $^{+}$ T cells.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | c-Jun overexpression mediates transcriptional but not epigenetic reprogramming of exhausted HA-28z CAR T cells. **a**, The \log_2 -transformed fold change in HA versus JUN-HA ATAC-seq demonstrating no significantly different peaks between conditions. **b**, Gene expression of 319 genes differentially expressed in JUN versus HA-28z CAR T cells (\log_2 -transformed fold change > 2 , $P_{\text{adj}} < 0.05$). Genes downregulated in JUN CAR T cells (blue) include exhaustion-associated genes such as *BATF3*, *GZMB*, *LAG3*, *JUNB* and *ENTPDI* (encoding CD39). Genes upregulated in JUN CAR T cells (red) include genes associated with naive and memory differentiation such as *IL7R*, *LEF1*, *SELL* (CD62L), *CD44*, and *KLF3*. **c**, Venn diagrams showing overlap of the 319 genes differentially expressed in JUN versus HA-28z and the top 200 genes distinguishing exhausted (HA) and healthy (CD19) CAR T cells from PC1 in Fig. 1e. **f**, Genes downregulated in JUN CAR T cells overlap with exhaustion-associated (HA, PC1-exhausted) genes, and genes upregulated in JUN CAR T cells overlap with genes associated with healthy memory differentiation

(CD19, PC1-healthy). **d**, DAVID bioinformatics analysis of transcription factor-binding sites within the 319 genes differentially expressed in JUN CAR T cells reveals that the top transcription factor binding motif belongs to the AP-1 family (269 out of 319 genes). **e**, Proposed mechanisms of c-Jun-mediated rescue of T cell exhaustion. AP-1i indicates an exhaustion-associated AP-1 complex. **f**, Immunoblot of total c-Jun and phosphorylated c-Jun (p-c-Jun^{S73}) in control, JUN-WT and JUN-AA HA-28z CAR T cells. **g**, Immunoblot analysis of c-Jun protein expression in control and indicated JUN-variant-expressing HA-28z CAR T cells in either soluble or chromatin-bound cellular lysate fractions. c-Jun variants with deletions in the C-terminal DNA binding and leucine zipper dimerization domains (basic, LeuZ and bZIP) cannot bind chromatin and do not rescue functional activity. **h**, Decrease in mRNA expression of *BATF*, *BATF3* and *JUNB* in JUN HA-28z CAR T cells compared with HA-28z cells ($n = 3$ donors, normalized to *CD19* mRNA). *P* values determined by ratio paired two-tailed *t*-test. See Supplementary Fig. 1 for gel source data.

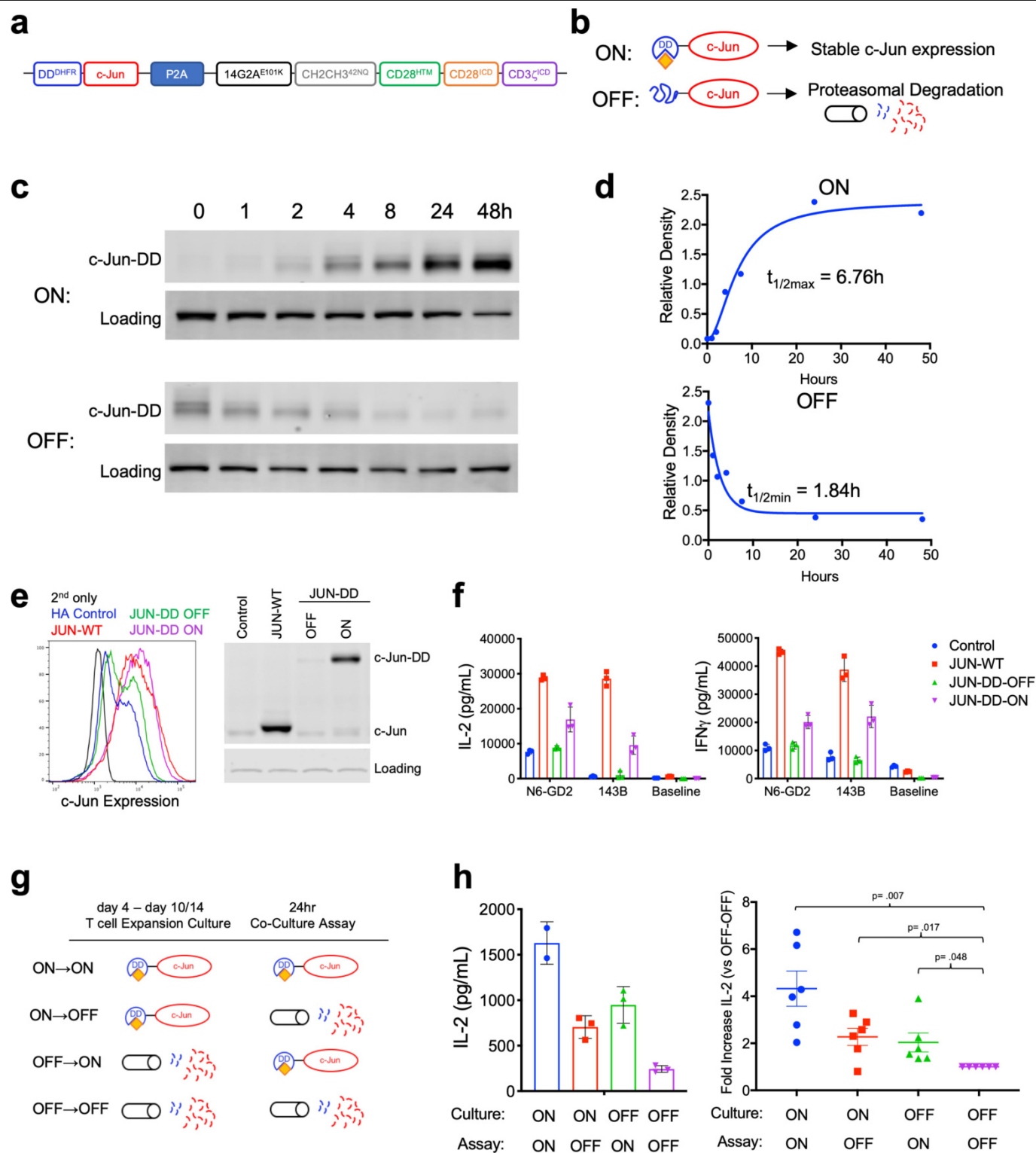


Extended Data Fig. 7 | See next page for caption.

Article

Extended Data Fig. 7 | c-Jun overexpression decreases chromatin binding and complexing of JunB–BATF–BATF3–AP-1 complexes. **a**, Immunoblot analysis for the indicated AP-1–bZIP and IRF family member proteins in control and JUN CD19-28z and HA-28z CAR T cells (day 10). **b**, Immunoblot analysis for the indicated AP-1–bZIP and IRF family member proteins in control and JUN HA-28z CAR T cells (day 10) in either soluble or chromatin-bound cellular lysate fractions. **c**, c-Jun overexpression decreases JunB–BATF and JunB–BATF3 complexes by immunoprecipitation–immunoblot analysis. Input (left), immunoprecipitation for c-Jun (middle), or JunB (right) in control or JUN HA-28z CAR T cells. Levels of IRF4 protein and complexes with c-Jun are unchanged. **d–h**, ChIP–seq analysis for c-Jun and IRF4. **d**, Motif enrichment in IRF4-bound (left) or c-Jun-bound (right) loci. **e**, IRF4 signal genome-wide. Data shown for each transduction at all IRF4-bound sites. The x and y axes show log-transformed normalized count signal in control and JUN-overexpression cells, respectively. **f**, IRF4 and c-Jun ChIP–seq genome tracks in JUN or control HA-28z CAR T cells. c-Jun ChIP with reference exogenous genome (ChIP–Rx; top), with x axis representing genomic position and y axis representing reference-adjusted reads per million (RRPM). IRF4 ChIP (bottom), with x axis representing genomic position and y-axis representing reads per million (RPM). Arrows indicate peaks with increased c-Jun binding in HA-28z JUN cells at IRF4-bound sites within genes previously described to be regulated by IRF4 or BATF (*TCF7*, *HAVCR2* and *HIF1A*)⁷. **g**, Overexpressed c-Jun is bound to IRF4-occupied sites in the genome. Enrichment plot of c-Jun ChIP–Rx signal (left) or IRF4 ChIP–seq

signal (right) in either JUN overexpression (red) or control (blue) HA-28z CAR T cells at all JUN-bound sites. The x axis shows distance from centre of JUN-bound site, and y axis shows average RRPM across replicates for c-Jun ChIP or average RPM across replicates for IRF4 ChIP. **h**, Venn diagram showing number of genes bound by IRF4 and/or c-Jun (*n* genes expressed/*n* genes bound). GSEA analysis with genes bound only by IRF4 (top) and genes bound by c-Jun and IRF4 (bottom), comparing levels of expression in JUN versus control HA-28z CAR T cells (normalized $P < 0.05$, FDR $< 25\%$). **i**, Immunoblot of indicated AP-1 and IRF protein in control or CRISPR-knockout HA-28z CAR T cells demonstrating productive knockout of target protein. **j**, IL-2 (top) and IFN γ (bottom) release in HA-28z CAR T cells with control or CRISPR-knockout of the indicated AP-1 or IRF4 gene after 24 h stimulation with Nalm6-GD2 or 143B target cells. Data are mean \pm s.d. of triplicate wells; representative of six independent experiments. Fold change across all experiments in Fig. 4e. In **k** and **l**, NSG mice were inoculated with 1×10^6 Nalm6-GD2 leukaemia cells via intravenous injection. A stress-test dose of 1×10^6 mock, HA-28z control, JUN-WT, JUN-AA or JUN- Δ bZIP HA-28z CAR⁺ T cells was given intravenously on day 7. **k**, Tumour progression was monitored using bioluminescent imaging. **l**, JUN-WT and JUN-AA HA-28z CAR T cells enhanced long-term survival, and control and JUN- Δ bZIP HA-28z CAR T cells were almost non-functional compared with mock untransduced T cells at this dose. Data are mean \pm s.e.m. of $n = 5$ mice per group. For gel source data see Supplementary Fig. 1.

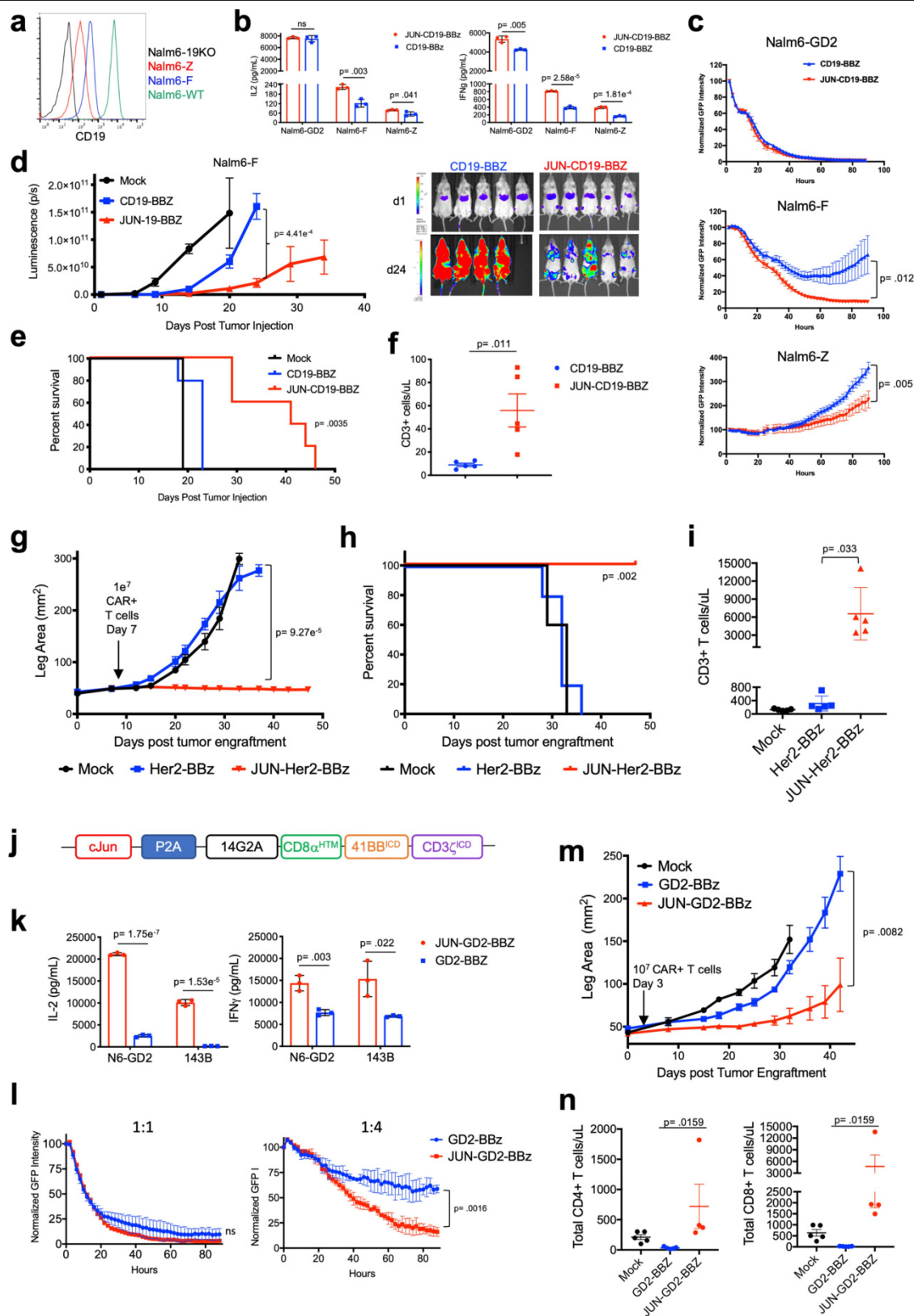


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Functional rescue of exhausted HA-28z CAR T cells requires the presence of c-Jun during both chronic and acute T cell stimulation.

a, Schematic of the destabilization domain (DD)-regulated JUN expression vector. **b**, Schematic of drug-induced stabilization of JUN-DD expression. Yellow diamond denotes trimethoprim (TMP)-stabilizing molecule. **c**, Kinetics of drug-induced c-Jun stability in JUN-DD CAR T cells as assessed by immunoblot. At time 0, 10 μ M TMP was either added to untreated cells (ON) or washed out of previously treated cells (OFF). Cells were removed from each condition at 1, 2, 4, 8, 24 and 48 h and prepared for immunoblot analysis of c-Jun expression. The observed band corresponds to the size of JUN-DD. **d**, Densitometry analysis was performed on the blots from **c** and normalized to loading control. Expression was plotted against time and first-order kinetics curves were fit to the data to determine $t_{1/2}$ for OFF and ON kinetics. **e**, Total c-Jun expression in control, JUN-WT and JUN-DD HA-28z CAR

T cells at day 10 by intracellular flow cytometry (left) and immunoblot (right). **f**, IL-2 (left) and IFN γ (right) production in control (blue), JUN-WT (red) or JUN-DD (OFF-green, ON-purple) modified HA-28z CAR T cells 24 h after stimulation with Nalm6-GD2 or I43B target cells, or media alone (baseline) at day 10. In **e** and **f**, OFF indicates without TMP, ON indicates T cells cultured in the presence of 10 μ M TMP from day 4 and during co-culture. In **g** and **h**, TMP was added either during T cell expansion (starting at day 4) or only during co-culture with tumour cells as indicated in **g**. For ON-to-OFF and OFF-to-ON conditions, TMP was removed or added 18 h before co-culture to ensure complete c-Jun degradation or stabilization, respectively, before antigen exposure. **h**, IL-2 expression in one representative donor (left, s.d. across triplicate wells) and fold increase in IL-2 (s.e.m. of $n = 6$ independent experiments representing three different donors, relative to OFF-OFF condition). *P* values determined by unpaired two-tailed *t*-tests. For gel source data, see Supplementary Fig. 1.

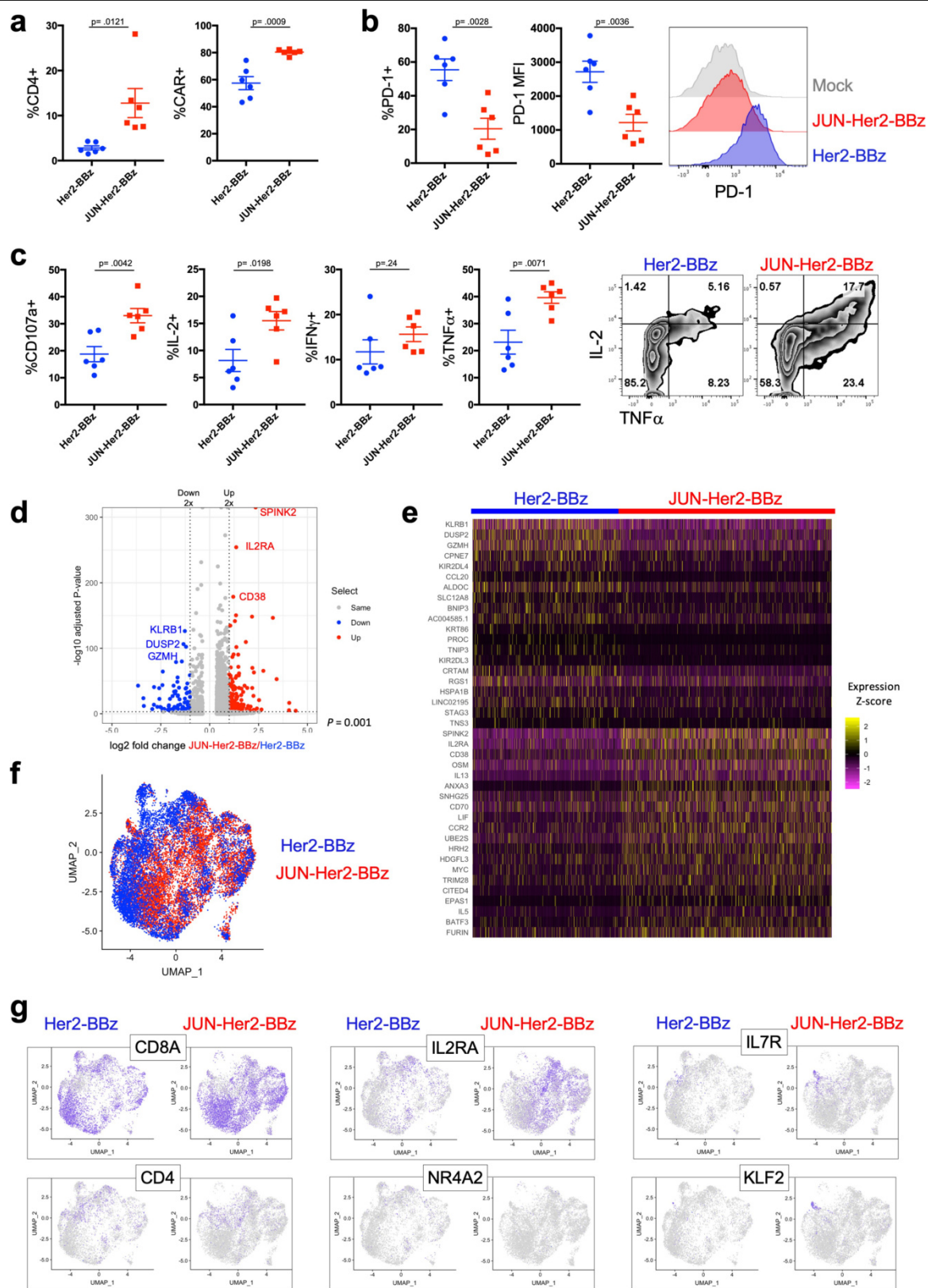


Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | c-Jun overexpression enhances CD19-BBz CAR T cells activity under suboptimal antigen stimulation and Her2 or GD2-BBz CAR T cell function in solid tumours. **a**, CD19 surface expression on parental Nalm6-WT (green), CD19-knockout Nalm6 (19KO, black), and two different CD19-low Nalm6 clones (F and Z) Nalm6-19KO plus CD19^{low}-F (F; blue), and Nalm6-19KO plus CD19^{low}-Z (red). **b**, IL-2 (left) and IFN γ (right) release after co-culture of control (blue) or JUN (red) CD19-BBz CAR T cells exposed to Nalm6-WT and Nalm6-19^{low} clones F and Z. **c**, JUN versus control CD19-BBz CAR T cell lysis of GFP⁺ Nalm6-WT (top), Nalm6-F (middle) or Nalm6-Z (bottom) target cells at a 1:2 effector:target cell ratio, demonstrating enhanced activity of JUN CAR T cells at low antigen density. Data in **b** and **c** denote mean \pm s.d. of triplicate wells; representative of four independent experiments. In **d–f**, NSG mice were inoculated with 1×10^5 Nalm6-19^{low} clone F leukaemia cells. On day 1, 3×10^6 control or JUN CD19-BBz CAR⁺ T cells or 3×10^6 mock-transduced T cells were transferred intravenously. **d**, Tumour growth was monitored by bioluminescent imaging. **e**, JUN expression significantly improved the long-term survival of CAR-treated mice. **f**, Mice receiving JUN-CD19-BBz CAR T cells display increased peripheral blood T cells on day 20. Data in **d–f** denote mean \pm s.e.m. of $n = 5$ mice per group; representative of three independent experiments. Long-term tumour-free survival is impeded in this model owing to outgrowth of CD19-negative disease. In **g–i**, NSG mice were inoculated with

1×10^6 143B osteosarcoma cells via intramuscular injection; then, 1×10^7 mock, Her2-BBz or JUN-Her2-BBz CAR T cells were given intravenously on day 7. **g**, Tumour growth was monitored by caliper measurements. **h**, Long-term survival. **i**, On day 20 after tumour implantation, peripheral blood T cells were quantified in mice treated as in **g**. Data are mean \pm s.e.m. of $n = 5$ mice per group; representative of two independent experiments. **j**, Vector schematic of JUN-GD2-BBz retroviral vector construct. **k**, IL-2 (left) and IFN γ (right) production in JUN-modified (red) or control (blue) GD2-BBz CAR T cells after 24 h stimulation with Nalm6-GD2 or 143B target cells. **l**, GD2-BBz CAR T cell lysis of GFP⁺ Nalm6-GD2 target cells at 1:1 (left) or 1:4 (right) effector:target cell ratio. Data in **k** and **l** denote mean \pm s.d. of triplicate wells; representative of four independent experiments. In **m** and **n**, NSG mice were inoculated with 0.5×10^6 143B-19 osteosarcoma cells via intramuscular injection; then, 1×10^7 mock, GD2-BBz or JUN-GD2-BBz CAR T cells were given intravenously on day 3. **m**, Tumour growth was monitored by caliper measurements. **n**, Peripheral blood CD4⁺ (left) or CD8⁺ (right) T cell counts at day 14 after tumour engraftment. Data are mean \pm s.e.m. of $n = 5$ mice per group; representative of two independent experiments although early deaths (unrelated to tumour size) precluded survival curves in both models. *P* values in **n** were determined by a Mann–Whitney test. All other *P* values determined by unpaired two-tailed *t*-tests. Survival curves were compared using the log-rank Mantel–Cox test.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | c-Jun overexpressing CAR TILs demonstrate increased activity in osteosarcoma xenograft tumours. Experimental design described in Fig. 6. **a–c**, Frequency (**a**), phenotype (**b**) and ex vivo functional activity (**c**) of CD4⁺ TILs from mice treated with Her2-BBz or JUN-Her2-BBz CAR T cells. **a**, Left, CD4⁺ levels as a proportion of total live tumour cells. Right, CAR⁺ as a frequency of total live CD4⁺. **b**, Percentage of PD-1⁺ (left) and PD-1 mean fluorescence intensity (MFI; middle) of total live CD4⁺ with representative flow histograms (right). Mock untransduced T cells were from spleens of tumour-bearing mice at the same time point. **c**, Frequency of indicated cytokine- or CD107a-producing cells after 5-h re-stimulation with Nalm6-Her2⁺ target cells. Gated on total, live CD4⁺ T cells (left) with representative contour plots (right). Data are mean \pm s.e.m. of $n = 6$ mice per group. Each data point represents an individual mouse. *P* values determined by unpaired two-tailed *t*-tests. In **d–g**, dissociated tumour cell suspensions were labelled and sorted by FACS analysis

to isolate live, human CD45⁺ TILs. Sorted cells from six mice per group were pooled and approximately 10,000 cells were processed for 3' single-cell RNA-seq on the 10X Genomics platform. **d**, Volcano plot showing results of differential expression analysis comparing JUN-Her2-BBz CAR T cells with control Her2-BBz CAR T cells. Top three upregulated and downregulated genes are highlighted. **e**, Heat map of the top 20 most significantly upregulated and downregulated genes. **f**, Uniform manifold approximation and projection (UMAP) embedding analysis showing JUN-Her2-BBz and control Her2-BBz CAR T cells overlaid. **g**, Expression of indicated transcripts in JUN-Her2-BBz or control Her2-BBz CAR T cells showing localization of CD4⁺ and CD8⁺ subsets, activation marker (*IL2RA*), exhaustion marker (*NR4A2*), and maintenance of a small memory-like population (*IL7RA*, *KLF2*) in JUN-overexpressing Her2-BBz CAR T cells within the solid osteosarcoma tumour microenvironment.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

FACSDiva ver8.0.1 (BD Biosciences) for flow cytometry data acquisition. LivingImage ver4.5 (Perkin Elmer) for IVIS Spectrum bioluminescent imaging acquisition. Image Studio Software ver3.1 (LI-COR) for image acquisition of western blots. Incucyte ZOOM ver2016B (Essen BioScience) for incucyte image acquisition (killing assay).

Data analysis

FlowJo ver 10.4.2 for analysis of flow cytometry. GSEA ver 2.0 (Broad Institute, UCSD) for gene set enrichment analysis. GraphPad Prism ver 7.0 (GraphPad) for biostatistical analysis and scientific graphing. Excel ver 16.14.1 (Microsoft) for spreadsheet management. SnapGene ver 4.2.1 (GSL Biotech LLC) for vector visualization and cloning. ImageJ ver 1.51J (NIH) for densitometry image analysis. LivingImage ver4.5 (Perkin Elmer) for IVIS Spectrum bioluminescent imaging analysis. Incucyte ZOOM ver2016B (Essen BioScience) for incucyte image analysis and quantification. DESeq2 (version 2.1.24) and ggplot2 (Version 3.2.1) package in R (ver 3.5.0) for PCA and plots of bulk RNA-seq data. For single cell RNA-seq analysis: CD19-28z vs. GD2-28z: Cell Ranger (version 1.2.0), R Studio (version 1.1.453), R (version 3.5.0), and packages: cellrangerRkit (version 2.0.0) dplyr (version 0.7.6) DESeq2 (version 1.20.0) GO.db (version 3.6.0) org.Hs.eg.db (version 3.6.0) gplots (version 3.0.1) pheatmap (version 1.0.10) and igraph (version 1.2.1). ggplot2 (version 3.0.0); Her2-BBz vs. JUN-Her2-BBz: Cell Ranger (version 3.0.2), R Studio (version 1.1.456), R (version 3.5.2), future (version 1.14.0), sctransform (version 0.2.0), gplots (version 3.0.1.1), Seurat (version 3.0.3.9035), dplyr (version 0.8.3), reshape2 (version 1.4.3), ggplot2 (version 3.2.1). For ATAC-seq analysis: SeqPurge, Bowtie2 (Johns Hopkins, ver2.1.0), Picard tools (Broad Institute, ver 1.79), And the following packages in R: MACS2 (ver 2.1.0.20150731), Bedtools, Motifmatchr, rtracklayer, DESeq2 (ver 1.20.0), and ChromVAR. ENCODE hg19 blacklist (UCSC genome browser), HOMER Motif Analysis ver 4.9 (UCSD). chromVAR ver 1.60.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing datasets generated in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession numbers: bulk RNA-seq: GSE136891, scRNA-seq CD19/GD2-28z: GSE136874, scRNA-seq Control/JUN-Her2-BBz TILs: GSE136805, ATAC-seq: GSE136796, ChIP-seq: GSE136853. Data Figures associated with these datasets are Figures 1, 2, and 6 and Extended Data Figures 1,2,3,4,6,7, and 10.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](#)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Our established animal models consistently yield low variance, and difference in the means between groups greater than 50%. Power calculation estimates determined that 5 animals per group was sufficient to determine statistical significance.
Data exclusions	No data were excluded from analysis
Replication	In vivo experiments were replicated at least 2 times with independent experiments (as stated in the figure legends) with similar results. All replication results provided the same trends and conclusions. As noted, long-term tumor-free survival in the Nalm6-GD2 model (Figure 5) was impacted due to outgrowth of GD2(-) clones in some experiments. In vitro experiments were typically run with triplicate technical replicates and were reproduced at least 3 times in independent experiments (as stated in the figure legends) with similar results.
Randomization	For in vivo tumor models, mice were randomized to ensure equal mean tumor burden before T cell transfer.
Blinding	Investigators were blinded during in vivo tumor measurement. Otherwise, fully informed data analysis was performed. Fully blinded experiments were not possible due to personnel availability to accommodate such situations.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials

Reagents (vector constructs) developed in house during this study are available upon request from the corresponding author.

Antibodies

Antibodies used

From BioLegend: CD4-APC-Cy7 (clone OKT4), CD8-PerCp-Cy5.5 (clone SK1), TIM3-BV510 (clone F38-2E2), CD39-FITC or APC-Cy7 (clone A1), CD95-PE (clone DX2), CD3-PacBlue (clone HIT3a), CD107a-BV607 (Clone H4A3), IL2-PECy7 Clone MQ1-17H12, IFNg-APC/Cy7 Clone 4S.B3, and TNFa-BV711 Clone Mab11.
 From eBioscience: PD1-PE-Cy7 (clone eBio J105), LAG3-PE (clone 3DS223H), CD45RO-PE-Cy7 (clone UCHL1), CD45-PerCp-Cy5.5 (clone H30),
 From BD: CD45RA-FITC or BV711 (clone HI100), CCR7-BV421 (clone 150503), CD122-BV510 (clone Mik-b3), CD62L-BV605 (clone DREG-56), CD4-BUV395 (clone SK3), CD8-BUV805 (clone SK1).
 1A7 obtained from NCI-Frederick, CD19-Idiotypic provided by Laurence Cooper.
 For western blot: Cell Signaling: GAPDH (D46CR), Histone-3 (1B1B2), c-Jun (60A8), P-c-JunSer73 (D47G9), JunB(C37F9), BATF(D7C5) and IRF4(4964). The BATF3 (AF7437) antibody was from R&D. The b-Actin (Clone AC-15)-HRP conjugated was from Sigma-Aldrich.

Validation

All antibodies used for flow cytometry were validated by the manufacturer directly in human peripheral blood mononuclear cells. In our laboratory, antibody-specific staining was compared to isotype and no staining control samples.
 CD4-APC-Cy7 (clone OKT4) was validated here <https://www.biolegend.com/en-us/products/apc-cyanine-7-anti-human-cd4-antibody-3658>
 CD8-PerCp-Cy5.5 (clone SK1) was validated here <https://www.biolegend.com/en-us/products/percp-cyanine5-5-anti-human-cd8-antibody-6389>
 TIM3-BV510 (clone F38-2E2) was validated here <https://www.biolegend.com/en-us/search-results/brilliant-violet-510-anti-human-cd366-tim-3-antibody-12009>
 CD39-FITC or APC-Cy7 (clone A1) was validated here <https://www.biolegend.com/en-us/products/fitc-anti-human-cd39-antibody-4363>
 CD95-PE (clone DX2) was validated here <https://www.biolegend.com/en-us/products/pe-anti-human-cd95-fas-antibody-643>
 CD3-PacBlue (clone HIT3a) was validated here <https://www.biolegend.com/ja-jp/search-results/pacific-blue-anti-human-cd3-antibody-6505>
 CD107a-BV607 (Clone H4A3) was validated here <https://www.biolegend.com/en-us/products/pe-anti-human-cd107a-lamp-1-antibody-4967>
 IL2-PECy7 Clone MQ1-17H12 was validated here <https://www.biolegend.com/en-gb/products/pe-cy7-anti-human-il-2-antibody-6513>
 IFNg-APC/Cy7 Clone 4S.B3 was validated here <https://www.biolegend.com/en-us/products/apc-anti-human-ifn-gamma-antibody-1012>
 TNFa-BV711 Clone Mab11 was validated here <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-human-tnf-alpha-antibody-9034>
 PD1-PE-Cy7 (clone eBio J105) was validated here <https://www.thermofisher.com/antibody/product/CD279-PD-1-Antibody-clone-eBioJ105-J105-Monoclonal/25-2799-42>
 LAG3-PE (clone 3DS223H) was validated here <https://www.thermofisher.com/antibody/product/CD223-LAG-3-Antibody-clone-3DS223H-Monoclonal/12-2239-42>
 CD45RO-PE-Cy7 (clone UCHL1) was validated here <https://www.thermofisher.com/antibody/product/CD45RO-Antibody-clone-UCHL1-Monoclonal/25-0457-42>
 CD45-PerCp-Cy5.5 (clone H30) was validated here <http://www.bdbiosciences.com/us/applications/research/stem-cell-research/cancer-research/human/percp-cy55-mouse-anti-human-cd45-hi30/p/564105>
 CD45RA-FITC or BV711 (clone HI100) was validated here <http://www.bdbiosciences.com/eu/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/human/bv711-mouse-anti-human-cd45ra-hi100/p/563733>
 CCR7-BV421 (clone 150503) was validated here <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-2-cells/surface-markers/human/bv421-mouse-anti-human-cd197-ccr7-150503/p/562555>
 CD122-BV510 (clone Mik-b3) was validated here <http://www.bdbiosciences.com/us/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/human/bv510-mouse-anti-human-cd122-mik-3/p/563093>
 CD62L-BV605 (clone DREG-56) was validated here <http://www.bdbiosciences.com/us/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/human/bv605-mouse-anti-human-cd62l-dreg-56/p/562720>
 CD4-BUV395 (clone SK3) was validated here <http://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/surface-markers/human/buv395-mouse-anti-human-cd4-sk3-also-known-as-leu3a/p/563550>

CD8-BUV805 (clone SK1) was validated here <http://wwwbdbiosciences.com/eu/reagents/research/antibodies-buffers/immunology-reagents/anti-human-antibodies/cell-surface-antigens/buv805-mouse-anti-human-cd8-sk1/p/612890>. This product is now discontinued

1A7 Anti-idiotype to detect GD2-CAR was first validated here Sen et al J Immunother 1998 and further utilized and validated in multiple publications (Long AH et al, Nat Med 2015; Alfonso M, J Immunol 2002; Mount CW, Nat Med 2018)

CD19-Idiotype provided by Laurence Cooper was validated here <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3585808/> and further utilized in multiple publications (Lee DW, The Lancet 2015; Long AH, Nat Med 2015; Kochenderfer JN, J Clin Oncol 2015; Fraietta JA, Nat Med 2018)

c-Jun (60A8), IRF4(4964), JunB(C37F9), BATF(D7C5), were validated by manufacturer (<https://www.cellsignal.com>) as well as by our lab using CRISPR-Cas9-edited human primary T cells followed by western blotting and/or flow cytometry.

BATF3 (AF7437) was validated by our lab using CRISPR-Cas9-edited human primary T cells followed by western blotting, as well as here https://www.rndsystems.com/products/human-batf3-antibody_af7437.

GAPDH (D46CR) was validated here <https://www.cellsignal.com/products/primary-antibodies/gapdh-d4c6r-mouse-mab/97166?N=4294956287&Ntt=gapdh&fromPage=plp>

b-Actin-HRP (clone AC-15) was validated here <https://www.sigmaaldrich.com/catalog/product/sigma/a3854?lang=en®ion=US>

P-c-JunSer73 (D47G9) was validated in our lab using CRISPR-Cas9-edited human primary T cells and overexpression vectors followed by western blotting. Further validation here <https://www.cellsignal.com/products/primary-antibodies/phospho-c-jun-ser73-d47g9-xp-rabbit-mab/3270>.

Histone 3 (1B1B2) was validated here <https://www.cellsignal.com/products/primary-antibodies/histone-h3-1b1b2-mouse-mab/14269>.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Nalm6-GL was originally provided by Steve Grupp (CHOP). EW8 and TC32 were provided by Lee Helman (NCI). The 143b line was provided by C. Khanna (NCI, NIH); Kelly from C. Thiele (NCI, NIH); 293GP line by the Surgery Branch (NCI). 293T cells were obtained from the American Type Culture Collection (ATCC, Manassas)
Authentication	STR DNA profiling of all cell lines is conducted by Genetica Cell Line testing once per year.
Mycoplasma contamination	Before using for in vivo experiments, cell lines are tested with MycoAlert detection kit (Lonza). All cell lines tested negative.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used in this study are included in the commonly misidentified cell lines registry.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	6-8 week old male or female NOD/SCID/IL2Rg (NSG) mice were used for all in vivo experiments.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Buffy coats from anonymous healthy (male and female human, ages 30-50) were purchased from the Stanford University Blood Bank.
Recruitment	Written informed consent was obtained from all healthy donors.

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	GEO Accession number: GSE136853
Files in database submission	<i>Provide a list of all files available in the database submission.</i>
Genome browser session (e.g. UCSC)	http://genome.ucsc.edu/s/snagara/Lynn%20et%20al%20enhanced%20CAR%2DT%20ChIP%2Dseq

Methodology

Replicates

ChIP-seq was performed in both JUN and Control (HA only) CAR-T cells in two biological replicates. Each replicate was a distinct viral transduction.

Sequencing depth

All ChIP-seq experiments were performed as paired-end sequencing at 2x75 bp length.

Unique reads:

ha-1-inp 89644144
ha-1-irf4 68466420
ha-1-jun 45203466
ha-2-inp 61388074
ha-2-irf4 39513462
jun-1-inp 61213392
jun-1-irf4 55079410
jun-1-jun 46053188
jun-2-inp 54899776
jun-2-irf4 50793788
jun-2-jun 42178818

Antibodies

JUN antibody: Active Motif, cat. no. 39309

IRF4 antibody: Abcam, cat. no. ab101168, lot. no. GR139704-21

Peak calling parameters

Reads were mapped using bowtie2 version 2.2.4 with the --very-sensitive flag. Peaks were called using MACS2 version 2.1.1.20160309 callpeak.

HA-1 IRF4 and JUN peaks were called over ha-1-inp.sort.rmdup.bam

HA-1 IRF4 peaks were called over ha-2-inp.sort.rmdup.bam

JUN-1 IRF4 and JUN peaks were called over jun-1-inp.sort.rmdup.bam

JUN-2 IRF4 and JUN peaks were called over jun-2-inp.sort.rmdup.bam

Data quality

ChIP peak enrichment was calculated using ENCODE QC parameters. Briefly, peak enrichment quality control was performed by downsampling bam files to 20 million reads using picardtools DownsampleSam and calling peaks again. ChIP and input coverage at these peaks was calculated using HOMER annotatePeaks.pl in 1 bp bins for 5 kb flanking each peak center.

Enrichment at each base pair was calculated as (ChIP coverage/input coverage) and overall peak enrichment was defined as the maximum enrichment in the 500 bp flanking the peak center.

Sample / Number of peaks called / Peak enrichment:

ha-1-irf4 / 52270 / 5.96

ha-1-jun / 1712 / 1.61

ha-2-irf4 / 55087 / 9.59

jun-1-irf4 / 43398 / 5.80

jun-1-jun / 7857 / 3.35

jun-2-irf4 / 43639 / 8.90

jun-2-jun / 9703 / 3.39

Software

Sequencing adaptors were trimmed using cutadapt version 1.11 and aligned to the hg19 reference genome using bowtie2 version 2.2.4 with the --very-sensitive flag. Aligned reads with quality less than 10 removed and remaining reads were sorted using samtools version 1.2. PCR duplicates were removed using picardtools version 1.128 MarkDuplicates.

For JUN ChIP-Rx, reads were first aligned to the hg19 genome and unmapped reads were mapped again to the dm6 Drosophila genome. Drosophila spike-in reads were filtered and deduplicated using the same parameters as described above. Sample H3K27me3 ChIP-seq were then normalized to reference-adjusted reads per million (RRPM), with normalization factor $N_d = 1,000,000 / (\text{number of unique Drosophila reads})$ used to make bigwig files for visualization.

For track visualization, each bam file was normalized to one million reads (rpm) or RRPM for JUN ChIP using bedtools version 2.19.1 genomecov -bg -split and then converting to bigwig format with bedGraphToBigWig. Tracks were displayed using the UCSC genome browser. Peaks were called using macs2 version 2.1.1.20160309 callpeak over input controls.

Correlation plots were generated by counting reads within merged peaks using bedtools multicov. Enrichment plots were generated using HOMER by counting reads in 5 bp bins surrounding peak centers.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Up to 1 million T cells from culture were washed with PBS + 2% FBS (FACS Buffer), labeled in 100uL FACS Buffer containing the relevant antibodies, and incubated at 4C in the dark for 20 minutes. Samples were washed 2X in 1mL FACS Buffer before running.

Instrument	BD LSRFortessa X-20
Software	FACSDiva for collection and FlowJo for analysis.
Cell population abundance	<i>Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.</i>
Gating strategy	All samples are gated on FSC/SSC lymphocyte populations, single cells (Using FSC-W/FSC-H and SSC-W/SSC-H), then either CD4 or CD8+. Frequency of positive gates is determined using an isotype or FMO control.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Developmental ROS individualizes organismal stress resistance and lifespan

<https://doi.org/10.1038/s41586-019-1814-y>

Received: 29 January 2019

Accepted: 18 October 2019

Published online: 4 December 2019

Daphne Bazopoulou¹, Daniela Knoefler¹, Yongxin Zheng^{2,3}, Kathrin Ulrich¹, Bryndon J. Oleson¹, Lihan Xie¹, Minwook Kim¹, Anke Kaufmann¹, Young-Tae Lee⁴, Yali Dou⁴, Yong Chen⁵, Shu Quan^{2,3} & Ursula Jakob^{1*}

A central aspect of aging research concerns the question of when individuality in lifespan arises¹. Here we show that a transient increase in reactive oxygen species (ROS), which occurs naturally during early development in a subpopulation of synchronized *Caenorhabditis elegans*, sets processes in motion that increase stress resistance, improve redox homeostasis and ultimately prolong lifespan in those animals. We find that these effects are linked to the global ROS-mediated decrease in developmental histone H3K4me3 levels. Studies in HeLa cells confirmed that global H3K4me3 levels are ROS-sensitive and that depletion of H3K4me3 levels increases stress resistance in mammalian cell cultures. In vitro studies identified SET1/MLL histone methyltransferases as redox sensitive units of the H3K4-trimethylating complex of proteins (COMPASS). Our findings implicate a link between early-life events, ROS-sensitive epigenetic marks, stress resistance and lifespan.

Genetic effects are estimated to account for only 10–25% of the observed differences in human lifespan¹. However, the remaining differences are not entirely attributable to environmental factors. Even when isogenic animals, such as *Caenorhabditis elegans*, are cultivated under identical environmental conditions, individual lifespans can vary by more than 50-fold. These results suggest that other, more stochastic factors account for variations in lifespan. Previous studies in *C. elegans* revealed that as early as day 1 of adulthood, subpopulations of longer-lived animals emerge². We therefore focused on the concept that specific fluctuating signals during development might differentially affect processes that determine lifespan. We investigated the idea that ROS³ might serve as early lifespan-determining modulators in *C. elegans*. This idea was built on *C. elegans* studies, which showed (1) that significant lifespan extension occurs following exposure to pharmacologically generated ROS in young adults⁴ or by altering mitochondrial activity during development⁵, (2) that exposure of nematodes to non-lethal concentrations of ROS leads to increased stress resistance and longevity, a phenomenon termed mitohormesis^{4,6} and (3) that individual larvae of a synchronized wild-type population exhibit large variations in endogenous ROS levels⁷.

Early-life ROS affect adult redox states

To investigate whether and how developmental ROS levels affect *C. elegans* later in life, we used wild-type N2 worms that ubiquitously express the integrated redox-sensing protein Grx1-roGFP2, which faithfully responds to the cellular ratio of oxidized and reduced glutathione (GSSG:GSH)⁸. Consistent with previous peroxide measurements⁷, L2 larvae revealed a significantly more oxidizing redox environment and substantially larger individual differences than young adults, which

exhibited a maximally reduced environment with smaller inter-individual differences (Extended Data Fig. 1a). With increasing age, the average redox state became more oxidizing and individual differences in redox status re-emerged. Subsequent analysis of about 16,000 age-synchronized L2 larvae using a reconfigured large particle BioSorter (Extended Data Fig. 1b) confirmed our microscopy studies and showed that the GSSG:GSH ratio varies widely among individuals (Fig. 1a). We sorted and binned L2 worms with redox states 2–3 standard deviations above (L2^{ox}) or below (L2^{red}) the mean population (L2^{mean}) (Fig. 1a, Extended Data Fig. 1c), and confirmed their different redox states by fluorescence microscopy (Fig. 1b, Extended Data Fig. 2a–d). Notably, L2^{ox} and L2^{red} worms did not differ significantly in size, reproductive activity, mitochondrial respiratory chain function or glycolytic flux (Extended Data Fig. 3a–f), excluding the possibility that more extreme early-life redox states affect development or other relevant physiological parameters. Subsequent redox analysis of sorted L2^{ox} and L2^{red} worms showed that all animals become similarly reduced in young adulthood and become more oxidized as they age (Fig. 1c). By day 7 of adulthood, however, the L2^{ox} worms were significantly more reduced than the L2^{red} worms. The trigger of the transient increase in GSSG:GSH ratios during early development and the mechanisms that cause the observed switch in endogenous redox states during adulthood are unknown. However, our results demonstrate that a synchronized population of *C. elegans* larvae contains subpopulations with redox environments that imprint information that becomes relevant later in life.

Early-life ROS extend lifespan

To investigate potential downstream effects of the observed variations in developmental redox levels, we compared stress resistance

¹Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI, USA. ²State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, China. ³Shanghai Collaborative Innovation Center for Biomanufacturing (SCICB), Shanghai, China. ⁴Department of Pathology, Michigan Medicine, Ann Arbor, MI, USA. ⁵State Key Laboratory of Molecular Biology, National Center for Protein Science Shanghai, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. *e-mail: ujakob@umich.edu

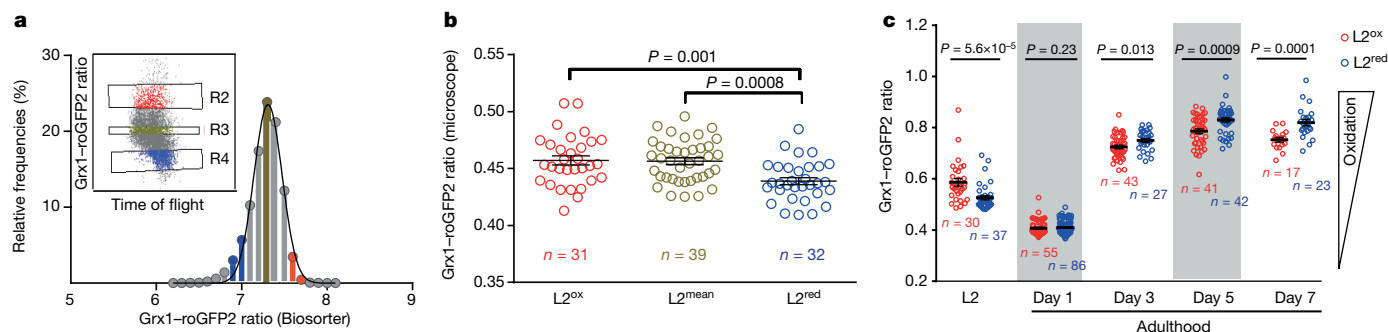


Fig. 1 | Endogenous redox state in an age-synchronized population of *C. elegans* larvae. **a**, Distribution of intracellular glutathione redox potential (Grx1-roGFP2 ratio) of an L2-staged N2[*grx1-roGFP2*] population.

L2 worms with Grx1-roGFP2 ratios between 2 and 3 standard deviations above ($L2^{ox}$; red line and inset, R2) or below ($L2^{red}$; blue line and inset, R4) the mean were sorted and compared with worms with mean Grx1-roGFP2 ratios ($L2^{mean}$; green line and inset, R3). $n = 15,599$ worms. **b**, Representative microscopy analysis of the Grx1-roGFP2 ratio of individual worms (dots) of the $L2^{ox}$, $L2^{mean}$

and $L2^{red}$ subpopulations. n , number of worms; one-way ANOVA with Tukey correction. The experiment was repeated four more times with similar results (see Extended Data Fig. 2a–d). **c**, Longitudinal analysis of the redox state. Sorted $L2^{ox}$ and $L2^{red}$ worms were cultivated at 20 °C and the Grx1-roGFP2 ratio of worms in each subpopulation was determined microscopically at the indicated time points. n , number of worms; two-sided Mann–Whitney U -test. In **b**, **c**, bars represent mean and error bars show s.e.m.

and lifespan of the sorted worms. We found that compared with $L2^{red}$ worms, $L2^{ox}$ worms were significantly more resistant to heat shock (Extended Data Fig. 3g), about 30% longer lived after the initial heat-shock treatment (Fig. 2a) and substantially longer lived when grown in the presence of oxidants such as paraquat (Fig. 2b) or juglone (Extended Data Table 1). Moreover, $L2^{ox}$ worms displayed an increase of up to 18% in median lifespan and a 1–4-day increase in maximal lifespan (Fig. 2c, Extended Data Fig. 2e–i, Extended Data Table 2). This clear correlation between increased GSSG:GSH ratios, stress resistance and lifespan suggested that a subpopulation of synchronized worms undergo naturally occurring hormesis during early development. Indeed, when we generated a more oxidizing environment in an entire population of worms by exposing them to 1 mM paraquat for 10 h during their L2 stage, the entire population became longer-lived (Fig. 2d with inset). Treatment of L2 worms with 10 mM *N*-acetylcysteine (NAC) for 10 h did not substantially alter the redox state of the population and had no significant lifespan effect. By conducting the same experiments with previously sorted subpopulations, however, we found that a 10-h NAC treatment decreased the lifespan of $L2^{ox}$ worms but not of the already reduced $L2^{red}$ worms, whereas a 10-h paraquat exposure increased the lifespan of $L2^{red}$ worms but not of the already oxidized $L2^{ox}$ worms (Fig. 2e, f, Extended Data Table 3). These results provide evidence that transient changes in the redox environment during early development are sufficient to positively affect the lifespan of *C. elegans*.

Early-life ROS reduce H3K4me3 levels

To gain insights into the mechanisms by which a transient increase in the cellular redox state during development might cause an increase in stress resistance and lifespan, we first conducted quantitative PCR with reverse transcription (RT-PCR) to examine changes in mRNA levels of commonly assessed heat-shock- (Fig. 3a) and oxidative stress-related genes (Extended Data Fig. 4a). Unexpectedly, $L2^{ox}$ and $L2^{red}$ worms did not significantly differ in the steady-state expression levels of any of these genes. However, on exposure to heat-shock conditions, $L2^{ox}$ worms showed a significantly increased capacity to upregulate heat-shock gene expression compared with $L2^{red}$ worms (Fig. 3a). Changes in transcript levels of heat-shock factor HSF-1 were not significantly different, suggesting that the transcriptional stimulation in $L2^{ox}$ worms is a result of either specific changes in HSF-1 activity, its subcellular localization or the accessibility of heat-shock promoter⁹.

Subsequent RNA-sequencing analysis from four independent large-scale sorting experiments identified 191 upregulated and 136

downregulated genes in $L2^{ox}$ worms compared with $L2^{red}$ worms (Extended Data Fig. 4b). We were unable to draw any clear link between the differentially expressed genes (DEGs) and previously identified sets of stress- or longevity-related genes (Extended Data Fig. 4c, Supplementary Table 1). However, 26 of the 191 upregulated genes in $L2^{ox}$ worms overlapped with a set of 101 genes previously shown to be upregulated in worms lacking the absent small homeodisc protein ASH-2¹⁰ (expected overlap if no correlation <1 ; $P = 1.7 \times 10^{-22}$, hypergeometric probability) (Fig. 3b, Supplementary Table 1). ASH-2 is a component of the highly conserved histone methylation complex COMPASS, which, together with a member of the SET1/MLL histone methyltransferase family (SET-2 in *C. elegans*) and other partner proteins, causes trimethylation of lysine 4 in histone H3 (H3K4me3)¹¹. H3K4me3 is primarily found at transcription start sites, where the modification is thought to mark and maintain transcriptionally active genes¹². Recent studies in *C. elegans* revealed that H3K4me3 marks associated with transcription start sites are set during early development and remain stable throughout life¹³. Indeed, analysis of published chromatin immunoprecipitation data revealed that about 25% of DEGs that we identified in $L2^{ox}$ worms associates with H3K4me3 marks that appear to be set during development (Extended Data Fig. 4d, e). Furthermore, we found a highly significant overlap between DEGs in $L2^{ox}$ worms and DEGs in strains lacking the H3K4me3 readers SET-9 or SET-26¹⁴, similar to the overlap between *ash-2* knockdown and *set-9* or *set-26* deletion strains (Extended Data Fig. 4f, g). On the basis of these results, we decided to analyse the global H3K4me3 abundance in $L2^{ox}$ and $L2^{red}$ worms by western blotting using antibodies against H3K4me3. We tested the subpopulations of seven independent sorting experiments and found a clear and highly reproducible reduction of more than 25% in global H3K4me3 levels in $L2^{ox}$ worms (Fig. 3c, Extended Data Fig. 5a). By contrast, other marks, such as H3K27ac or H3K27me3, were not significantly different between the two subpopulations (Extended Data Fig. 5b, c). These results strongly suggest a link between endogenous ROS levels, H3K4 trimethylation levels and gene regulation.

A redox-sensitive histone mark

RNA-sequencing analysis of $L2^{ox}$ and $L2^{red}$ worms did not reveal any transcriptional changes in components of the COMPASS complex. This result suggested that the activity rather than the level of the H3K4me3 complex is affected by the redox environment. To directly test whether the H3K4me3 machinery is ROS-sensitive, we attempted to purify the *C. elegans* proteins for in vitro methylation assays. However,

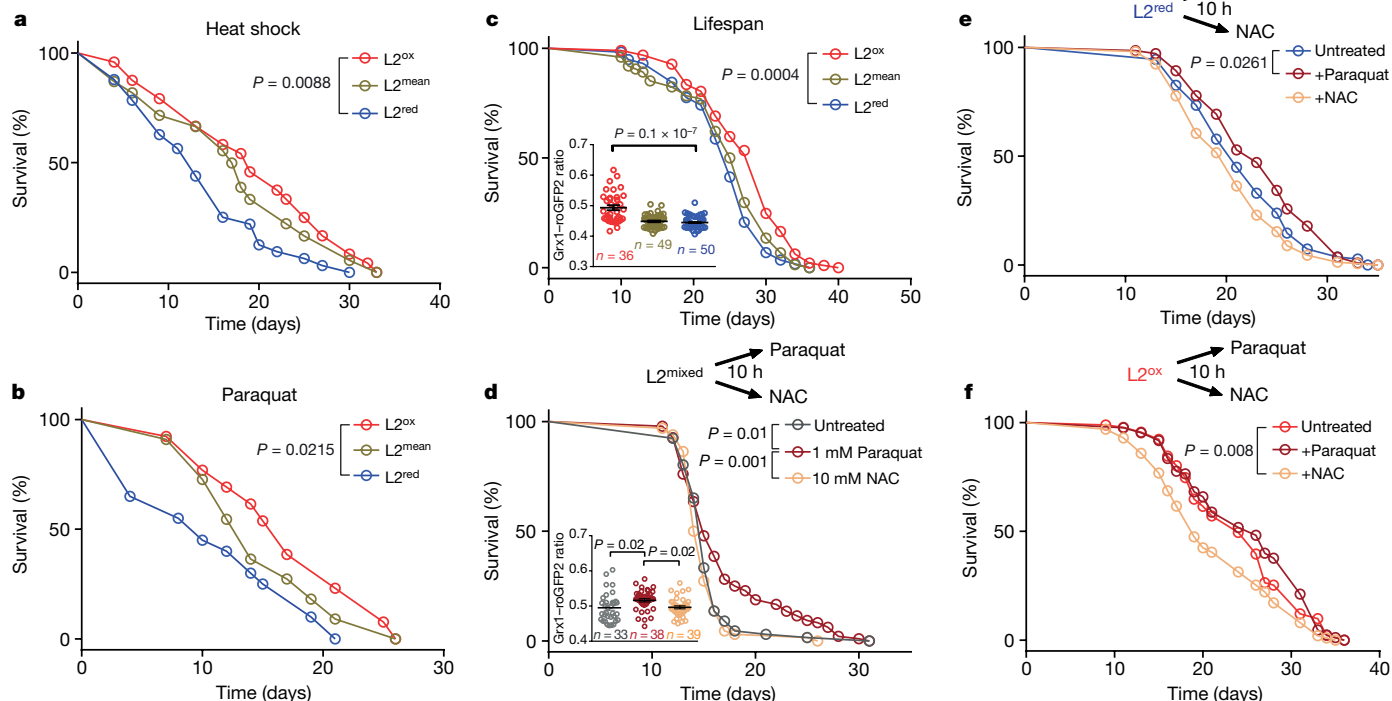


Fig. 2 | Oxidized L2 subpopulations show increased stress resistance and longer lifespan. Experiments were performed with *N2jrls2[Prpl-17::Grx1-roGFP2]* worms sorted into $L2^{ox}$, $L2^{mean}$ and $L2^{red}$ subpopulations. **a**, Representative survival curves of sorted worms that survived heat-shock treatment. **b**, Representative survival curves of sorted worms cultivated on NGM plates supplemented with 2 mM paraquat. **c**, Representative survival curves of sorted worms. See Extended Data Fig. 2e–i for repetitions. Inset, Grx1-roGFP2 ratio of individual worms (dots) after sorting. n , number of worms; two-sided unpaired t -test. **d**, Representative survival curves of a non-

sorted (mixed) worm population either untreated or treated at the L2 stage with 1 mM paraquat or 10 mM NAC for 10 h. Inset, Grx1-roGFP2 ratio of individual worms (dots) after treatment. n , number of worms; one-way ANOVA with Tukey correction. Data in insets are mean \pm s.e.m. **e**, **f**, Representative survival curves of an $L2^{red}$ (**e**) or $L2^{ox}$ (**f**) subpopulation, either untreated or after a 10-h treatment with 1 mM paraquat or 10 mM NAC. The specific sorting events, number of individuals, repetitions and statistical analysis (log-rank) for each of the datasets shown in this figure are presented in Extended Data Tables 1–3.

this turned out to be impossible owing to their instability. By contrast, the mammalian members of the COMPASS have been previously used in H3K4 methylation assays *in vitro*¹⁵. To determine whether H3K4 trimethylation is a redox-sensitive process in mammalian cells, we subjected HeLa cells to non-lethal H_2O_2 treatment and investigated global H3K4me3 levels. Consistent with our findings in *C. elegans*, we observed a global decrease of about 30% in H3K4me3 levels within 30 min of peroxide treatment (Fig. 3d, Extended Data Fig. 5d), without detectable changes in steady-state levels of two of the main COMPASS components (Extended Data Fig. 5e, f). We subsequently purified the mammalian core COMPASS proteins—that is, the SET domain of MLL1, which catalyses lysine-directed histone methylation¹⁶, ASH2L, WDR5 (human homologue of *C. elegans* WDR-5.1) and RBBP5. We then treated the individual proteins with peroxide for 30 min, removed the oxidant, recombined the proteins and tested them for *in vitro* histone methylation activity. Notably, among these proteins, only the SET-domain of MLL1 appeared to be reproducibly peroxide-sensitive (Fig. 3e, Extended Data Fig. 5g, h). Incubation of the peroxide-inactivated SET domain with thiol-reducing dithiothreitol (DTT) restored the original activity, strongly suggesting that thiol oxidation is responsible for the reversible inactivation (Fig. 3e, Extended Data Fig. 5g, h). Analysis of the SET domain of other SET1/MLL-family members¹⁷, including SET1A, the most closely related human homologue of *C. elegans* SET-2¹⁸ (Extended Data Fig. 5i), SET1B (Extended Data Fig. 5j) and a version of MLL1 lacking the cysteine-containing GST-tag used for purification (Extended Data Fig. 5h) demonstrated that sensitivity towards peroxide is a universal feature for SET1/MLL family members. To investigate which of the seven cysteines in the SET domain of MLL1 might be sensitive to

reversible thiol oxidation, we conducted direct (Extended Data Fig. 5k, l) and reverse thiol trapping (Fig. 3f) on oxidized and reduced MLL1 SET domains using the 500-Da thiol-reactive compound 4-acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid (AMS), followed by SDS-PAGE¹⁹. Analysis of the migration behaviour of thiol-trapped oxidized versus reduced MLL1 SET domain suggested that peroxide treatment leads to the formation of two intramolecular disulfide bonds. Subsequent mass spectrometry analysis of *in vitro*-modified cysteines confirmed these results and revealed that at least four of the five absolutely conserved cysteines in the SET domain are highly sensitive to oxidation (Extended Data Fig. 5m–o). To our knowledge, H3K4me3 is the only histone methylation mark and MLL1 is the only histone methyltransferase that is known to be posttranslationally redox-regulated.

H3K4me3 modulates stress resistance

Our studies raised the possibility that redox-mediated inactivation of the lone SET1/MLL-homologue in the oxidized subpopulation of *C. elegans* (that is, SET-2) leads to a reduction in global H3K4me3 levels, which causes increased stress resistance and longevity. This theory was supported by recent studies showing that deleting or knocking down components of COMPASS in *C. elegans* increases lifespan¹⁰. Indeed, strains with either deleted (that is, *set-2* or *wdr-5.1*) or depleted (that is, *ash-2*) members of COMPASS were significantly more heat stress resistant than respective control strains or a strain deficient in the H3K4me3-demethylase RBR-2 (Fig. 3g). In addition, and comparable to results obtained with $L2^{ox}$ worms, we found that worms deficient in H3K4me3 (Extended Data Fig. 6a–c) showed a substantially increased

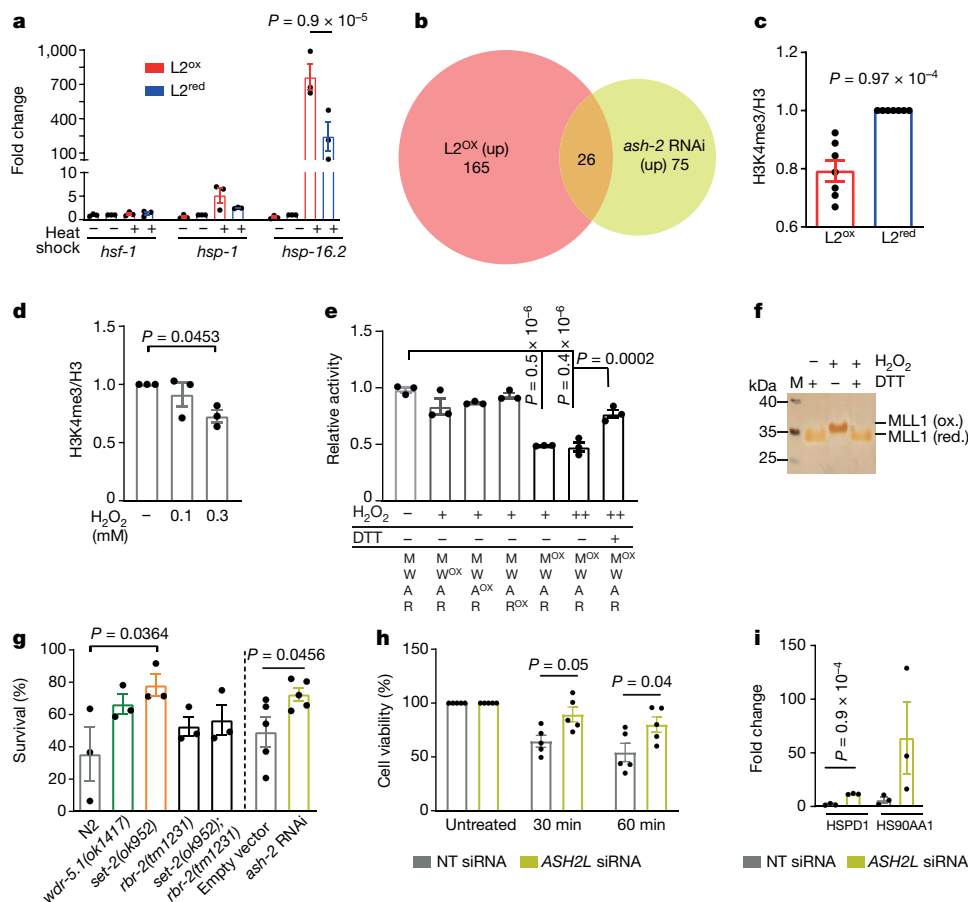


Fig. 3 | H3K4me3 is a redox-sensitive histone modification involved in stress gene expression and resistance. **a**, Transcripts of heat-shock-associated genes in sorted subpopulations before and after heat shock. $n = 3$ independent sorting experiments; two-way ANOVA with Tukey correction. **b**, Venn diagram of upregulated genes in $L2^{ox}$ and *ash-2* RNAi worms¹⁰ (complete list in Supplementary Table 1). **c**, Quantification of global H3K4me3 levels in $L2^{ox}$ and $L2^{red}$ worms. $n = 7$ independent sorting experiments; unpaired two-sided *t*-test. **d**, Quantification of H3K4me3 levels in HeLa cells before and after H_2O_2 treatment. $n = 3$ independent experiments; one-way ANOVA with Dunnett correction. **e**, In vitro histone methyltransferase assays with the purified core COMPASS proteins MLL1 (M), WDR5 (W), ASH2L (A) and RBBP5 (R); OX in superscript indicates that the protein is pre-treated with 1 mM (+) or 2 mM (++) H_2O_2 for 30 min before the activity assay. DTT was added after the H_2O_2 treatment as indicated. $n = 3$ independent experiments; one-way ANOVA with

Sidak correction. **f**, Reverse thiol trapping of oxidized and reduced MLL1 SET domain. A 500-Da mass increase per oxidized thiol can be detected on non-reducing SDS-PAGE. M, protein marker. **g**, Heat-shock survival of wild-type, *wdr-5.1*, *set-2*, *rbr-2* and *set-2/rbr-2* worms after 48 h ($n = 3$ independent experiments; one-way ANOVA with Dunnett correction) or *N2/jrIs2[Prpl-17::Grx1-roGFP2]* worms treated with *ash-2* or control RNAi after 24 h ($n = 5$ independent experiments; two-sided unpaired *t*-test). **h**, Heat-shock survival of HeLa cells treated with *ASH2L* siRNA. $n = 5$ independent experiments; two-way ANOVA with Tukey correction. **i**, Transcript levels of heat-shock genes after 30-min heat stress treatment of HeLa cells treated with *ASH2L* siRNA. NT, non-targeting. $n = 3$ independent experiments; two-sided unpaired *t*-test. Data in **a**, **c–e** and **g–i** are mean \pm s.e.m. For blot and gel source images, see Supplementary Figs. 1–3.

transcriptional response after heat shock (Extended Data Fig. 6d, e). We obtained very similar results with HeLa cells depleted of ASH2L by short interfering RNA (siRNA) treatment (Extended Data Fig. 6f), which were more resistant to heat stress (Fig. 3h) and demonstrated an augmented transcriptional response to heat stress compared with control siRNA-treated cells (Fig. 3i). These results strongly suggested that the downstream effects of H3Kme3 depletion on stress resistance and longevity are conserved.

Finally, to test whether downregulation of H3K4me3 levels is sufficient to increase heat-shock resistance and lifespan in $L2^{ox}$ worms, we generated *ash-2* or *set-2* RNA interference (RNAi)-mediated knockdown worms expressing the Grx1-roGFP2 redox sensor protein and sorted the synchronized $L2$ population as before. We detected no significant difference in the relative distribution or range of GSSG:GSH ratios between knockdown and control RNAi worms (Extended Data Fig. 6g). However, the sorted $L2^{ox}$ and $L2^{red}$ subpopulations of *ash-2* and *set-2* RNAi worms no longer exhibited any difference in heat-shock sensitivity (Fig. 4a, b) or lifespan (Fig. 4c, d, Extended Data Tables 2 and 4). Similarly, there was

no life-prolonging effect when we treated *ash-2* or *set-2* RNAi worms with paraquat for 10 h at the $L2$ larval state (Extended Data Fig. 6h, Extended Data Table 4). These results imply that downregulation of H3K4me3 levels is both necessary and sufficient to increase heat-shock resistance and lifespan in the oxidized subpopulation of worms.

Conclusions

In 2010, Cynthia Kenyon suggested that a stochastic event might flip an epigenetic switch or set in motion a chain of events that promotes ageing²⁰. Our studies have revealed that variations in endogenous ROS during development—potentially caused by locally different growth conditions—contribute to the variation in lifespan observed in synchronized populations of *C. elegans*. Animals that accumulate high levels of ROS during development apparently undergo an endogenous hormesis event, which, as previously observed with exogenous ROS treatment²¹, increases stress resistance and lifespan. This might serve as a bet-hedging strategy to provide subpopulations of worms with

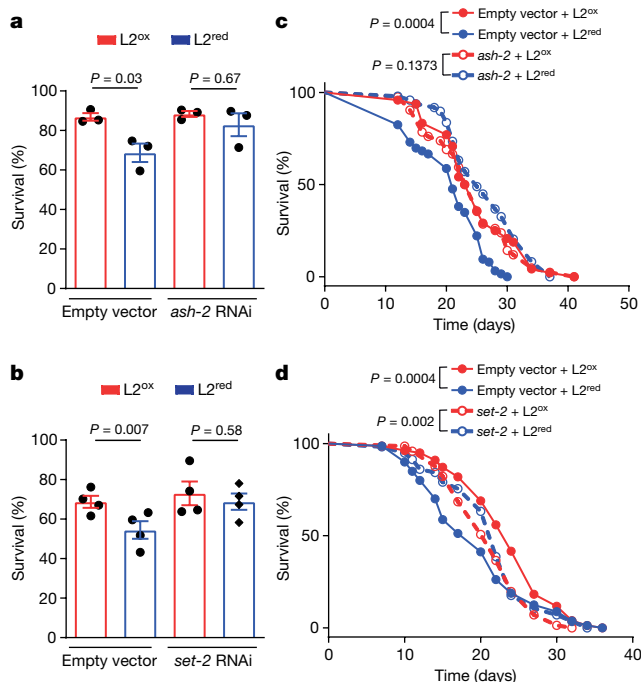


Fig. 4 | An intrinsically oxidizing environment confers increased stress resistance via downregulation of global H3K4me3 levels. a, b, Survival of *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with *ash-2* (a) or *set-2* (b) RNAi, sorted into *L2^{ox}* and *L2^{red}* and counted 24 h after heat shock. $n = 3$ (a) and $n = 4$ (b) independent experiments; two-way ANOVA with Tukey correction. Data are mean \pm s.e.m. **c, d,** Representative survival curves of *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with *ash-2* (c) or *set-2* (d) RNAi and sorted into *L2^{ox}* and *L2^{red}*. For *n* numbers, repetitions and statistics (log-rank) in c and d, see Extended Data Tables 2 and 4.

improved survival during stress. Our studies reveal the underlying mechanism of ROS-mediated hormesis by demonstrating that global H3K4me3 levels are redox-sensitive and decrease in response to oxidative stress. On the basis of the findings that decreased global H3K4me3 levels increase stress resistance and *C. elegans* lifespan¹⁰, we postulate that we have identified a stochastic event, the epigenetic switch and the chain of events that are set in motion during early development to increase lifespan.

Recent studies in *C. elegans* demonstrated that H3K4me3 marks within gene bodies are set during adulthood and change with age¹³. By contrast, H3K4me3 enrichment at transcription start sites—which is thought to provide a ‘memory’ of actively transcribed genes—is set during development and remains stable throughout the lifespan¹³. This result probably explains how transient redox-mediated changes in H3K4me3 levels during development are sufficient to exert long-lasting effects despite the marked changes in the redox environment during adulthood. Our finding that organisms with lower levels of H3K4me3 show an increased transcriptional response to stress conditions is initially counterintuitive, given that H3K4me3 is widely considered an activating mark¹². However, our results are fully consistent with yeast studies, which have shown that reduction of H3K4me3 levels causes substantially more robust gene expression changes upon stress treatment²². The extent to which this increase in transcriptional capacity of stress-related genes is linked to the observed lifespan extension, however, remains to be determined. A substantial number of genes involved in lipid metabolism were found to be downregulated in *L2^{ox}* worms (Extended Data Fig. 4c, Supplementary Table 1). This is notable, as increased lipid storage and altered lipid signalling have previously been linked to increased lifespans of a variety of different organisms²³

and found to have a role in the extension of lifespan of worms globally deficient in H3K4me3 levels²⁴. Further investigation is needed to reveal how the transient downregulation of H3K4me3 levels selectively during development can elicit similarly profound lifespan-altering effects. Our ability to change the lifespan of an entire population with a simple 10-h exposure to ROS during development suggests that we have identified a time window and a mechanism that helps to individualize lifespan in animals. This study provides a foundation for future work in mammals, in which very early and transient metabolic events in life seem to have equally profound impacts on lifespan²⁵.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1814-y>.

- Finch, C. E. & Tanzi, R. E. Genetics of aging. *Science* **278**, 407–411 (1997).
- Rea, S. L., Wu, D., Cypser, J. R., Vaupel, J. W. & Johnson, T. E. A stress-sensitive reporter predicts longevity in isogenic populations of *Caenorhabditis elegans*. *Nat. Genet.* **37**, 894–898 (2005).
- Holmström, K. M. & Finkel, T. Cellular mechanisms and physiological consequences of redox-dependent signalling. *Nat. Rev. Mol. Cell. Biol.* **15**, 411–421 (2014).
- Schulz, T. J. et al. Glucose restriction extends *Caenorhabditis elegans* life span by inducing mitochondrial respiration and increasing oxidative stress. *Cell Metab.* **6**, 280–293 (2007).
- Dillin, A. et al. Rates of behavior and aging specified by mitochondrial function during development. *Science* **298**, 2398–2401 (2002).
- Ristow, M. & Schmeisser, S. Extending life span by increasing oxidative stress. *Free Radic. Biol. Med.* **51**, 327–336 (2011).
- Knoefler, D. et al. Quantitative in vivo redox sensors uncover oxidative stress as an early event in life. *Mol. Cell* **47**, 767–776 (2012).
- Gutsch, M. et al. Real-time imaging of the intracellular glutathione redox potential. *Nat. Methods* **5**, 553–559 (2008).
- Labbadia, J. & Morimoto, R. I. Repression of the heat shock response is a programmed event at the onset of reproduction. *Mol. Cell* **59**, 639–650 (2015).
- Greer, E. L. et al. Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* **466**, 383–387 (2010).
- Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
- Pu, M. et al. Unique patterns of trimethylation of histone H3 lysine 4 are prone to changes during aging in *Caenorhabditis elegans* somatic cells. *PLoS Genet.* **14**, e1007466 (2018).
- Wang, W. et al. SET-9 and SET-26 are H3K4me3 readers and play critical roles in germline development and longevity. *eLife* **7**, e34970 (2018).
- Dou, Y. et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* **13**, 713–719 (2006).
- Southall, S. M., Wong, P. S., Odho, Z., Roe, S. M. & Wilson, J. R. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol. Cell* **33**, 181–191 (2009).
- Shilatifard, A. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.* **20**, 341–348 (2008).
- Li, T. & Kelly, W. G. A role for Set1/MLL-related components in epigenetic regulation of the *Caenorhabditis elegans* germ line. *PLoS Genet.* **7**, e1001349 (2011).
- Leichert, L. I. et al. Quantifying changes in the thiol redox proteome upon oxidative stress in vivo. *Proc. Natl Acad. Sci. USA* **105**, 8197–8202 (2008).
- Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
- Yang, W. & Hekimi, S. A mitochondrial superoxide signal triggers increased longevity in *Caenorhabditis elegans*. *PLoS Biol.* **8**, e1000556 (2010).
- Weiner, A. et al. Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol.* **10**, e1001369 (2012).
- Hansen, M., Flatt, T. & Aguilar, H. Reproduction, fat metabolism, and life span: what is the connection? *Cell Metab.* **17**, 10–19 (2013).
- Han, S. et al. Mono-unsaturated fatty acids link H3K4me3 modifiers to *C. elegans* lifespan. *Nature* **544**, 185–190 (2017).
- Sun, L., Sadighi Akha, A. A., Miller, R. A. & Harper, J. M. Life-span extension in mice by preweaning food restriction and by methionine restriction in middle age. *J. Gerontol. A* **64A**, 711–722 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments, unless otherwise noted, and outcome assessment.

C. elegans strains, maintenance and lifespan assays

The following *C. elegans* strains were used in this study: PB020: N2jrls2[Prpl-17::Grx-1-roGFP2], N2: Wild-type Bristol isolate, RB1304: *wdr-5.1(ok1417)*, RB1025: *set-2(ok952)*, ZR1: *rbr-2(tm1231)* and ABR9: *set-2(ok952);rbr-2(tm1231)*. Unless stated otherwise, worms were cultured at 20 °C. Standard procedures were followed for *C. elegans* strain maintenance²⁶. Synchronization was performed using alkaline hypochlorite solution; eggs were allowed to hatch by overnight incubation in M9 medium during gentle shaking. Newly hatched, arrested L1 larvae were transferred onto standard nematode growth medium (NGM) plates seeded with live *E. coli* OP50. Lifespan studies were performed at 20 °C in the presence of fluorodeoxyuridine. Survival was scored every 2 days, and worms were censored if they crawled off the plate, hatched inside or lost vulva integrity during reproduction. The first day of adulthood was set as $t = 0$. Lifespan of unsorted worm populations were performed in a lifespan machine according to²⁷. Survival plots were generated using GraphPad Prism. Lifespan data were analysed for statistical significance with log-rank (Mantel–Cox) or Gehan–Breslow–Wilcoxon test.

Reconfiguration of BioSorter for ratiometric sorting

Lasers of wavelength 405 and 488 nm were used to excite the Grx1-roGFP2 sensor protein. Since the protein possesses a single emission maximum (~520 nm), the two lasers in the BioSorter (Union Biometrica) were realigned to sequentially illuminate single L2-staged worms as they passed through the flow cell, without emitting overlapping signals. This enabled collection of signals from 405- and 488-nm lasers separately, from two photon multipliers tubes. As a result, data were displayed as two groups of peaks (Extended Data Fig. 1b). Using the partial profiling feature (pp) of the FlowPilot-Pro software, we mapped the peaks corresponding to each laser that trace the fluorescent intensity and extinction signals. The extinction signal from the 488-nm laser was used to initially gate worms at the L2 stage larva (R1 gate, see Extended Data Fig. 1c). Oxidized, mean and reduced L2 worms were sorted from R2, R3 and R4 gates respectively, based on the peak 405 and 488-nm fluorescent intensities (inset in Fig. 1a).

Microscopy

Worms were mounted on objective slides using 4 µl thermoreversible CyGel (BioStatus; Fisher Scientific) and 2 µl of 50 mM levamisole for immobilization. Fluorescence and DIC images were acquired with an upright microscope equipped with a Photometrics Coolsnap HQ2 cooled CCD camera, a UPlan S-Apo 20× objective (NA 0.75) and a X-CITE exacte light source equipped with a closed feedback loop. For Grx1-roGFP2 fluorescence, an external filter wheel was used with excitation filters 420/40×, 500/20×, dual bandpass dichroic T515LPXR and a single emission filter 535/30×. Image analysis was performed in Metamorph (Molecular Devices) using a custom script. In brief, an intensity threshold was chosen by the user. Pixels above this threshold constitute regions of interest. Regions with very high signal in any channel (for example, fluorescent particles) were identified by applying an over-saturation threshold and excluded from regions of interest. Mean ratiometric values (excitation 420 nm–emission 535 nm/excitation 500 nm–emission 535 nm) of regions of interest were calculated after subtraction of background. Acquisition parameters were kept identical across all samples. For body-length measurements, worms were measured from the nose to the tail tip and analysis was performed with ImageJ.

Brood size

L4-staged worms were transferred onto NGM plates and incubated at 15 °C. The parental animals were transferred daily to individual NGM

plates until the end of the reproductive period. The progeny of each animal was counted at the L2 or L3 stage.

Cellular respiration

Real-time oxygen consumption rates and extracellular acidification rate were measured with a Seahorse XF⁹⁶ Analyzer (Seahorse Bioscience) as described²⁸. In brief, 100 L2-staged worms were sorted directly into individual wells of 96-well Seahorse utility plates at a final volume of 200 µl of 10% M9. Acute effects of pharmacological inhibitors carbonyl cyanide-4(*p*-trifluor-methoxy)phenylhydrazone (FCCP), an accelerator of the electron transport chain (ETC) and sodium azide (NaN₃), a complex IV and V inhibitor, were evaluated by injecting them during the run at final concentrations of 20 µM and 40 mM respectively.

Heat-shock treatment

Heat shock was performed on solid OP50-seeded NGM plates wrapped in parafilm and submerged in a pre-heated water bath. For thermotolerance assays, worms were heat-shocked for 45 min at 38 °C. Survival was scored after 24 h and then until the death of the last worm by absence of touch response or pharyngeal pumping. For transcriptional response assays, worms were heat-shocked for 30 min at 35 °C. After 1 h recovery at 20 °C, worms were collected and snap-frozen in liquid nitrogen. All heat-shock treatments were applied to worms at the L2 stage.

Treatments with NAC and paraquat

For survival assays, worms were cultivated on solid OP50-seeded NGM plates, supplemented with the indicated concentrations of NAC or paraquat. Survival was determined by absence of touch response or pharyngeal pumping. For transient exposure to compounds, worms were transferred into M9-media supplemented with OP50 and the indicated concentrations of NAC or paraquat for 10 h. Worms were harvested, washed three times with M9 and transferred onto regular OP50-seeded NGM plates.

RNA extraction and real-time quantitative PCR

Worms at the L2 stage (3,000–5,000; whole population or after sorting) were grounded in Trizol reagent (Life Technologies) with sea sand and a pestle. After filtering of the sand, samples were vigorously shaken with chloroform, allowed to stand for 3 min at room temperature, and then centrifuged at 16,000g at 4 °C. The aqueous phase was then collected and RNA was purified using QIAGEN RNeasy RNA extraction columns according to the manufacturers' recommendations. HeLa cells were directly lysed in the culture dish by adding Trizol, as per manufacturers' recommendations. For RNA isolation, after addition of ethanol, the lysate was loaded onto QIAGEN RNeasy RNA extraction columns. cDNA synthesis was performed using PrimeScript 1st strand cDNA Synthesis Kit (Takara) and real-time quantitative PCR was performed using Radiant Green Lo-ROX qPCR Kit (Alkali Scientific), as per manufacturers' recommendations, in an Eppendorf Mastercycler epgradient S realplex² detection system. Relative expression was calculated from Cycle threshold values using the $2^{-\Delta\Delta C_t}$ method and the expression of genes of interest were normalized to housekeeping genes (*cdc-42*, *pmp-3*, *panactin*) and/or spiked-in luciferase (10 pg ml⁻¹ Trizol).

Primers used were *cdc-2*: 5'-AGCCATTCTGGCCGCTCTCG-3' and 5'-GCAACCGCTTCTCGTTGGC-3'; *pmp-3*: 5'-TTTGTGTCAATTGGTCATCG-3' and 5'-CTGTGTCAATGTCGTGAAGG-3'; *panactin*: 5'-TCGGTATGGGACAGAAGGAC-3' and 5'-CATCCAGTTGGTGACGATA-3'; *sod-1*: 5'-AAAATGTGGAACCGTGCTG-3' and 5'-TGAACGTGGAATCCATGAA-3'; *sod-2*: 5'-GATTGGAGCCTGTAATCAGTC-3' and 5'-GAAGAGCGATAGCTTCTTTGAC-3'; *sod-3*: 5'-CACTATTAAGCGCGACTTTCGG-3' and 5'-CAATATCCCAACCATCCCCAG-3'; *ctl-2*: 5'-ATCCCAACATGATCTTTGA-3' and 5'-TGAGATTCTTCACTGGTTG-3'; *prdx-2*: 5'-CGACTCTGTCTTCTCTCAC-3' and 5'-GAAGATCATTGATGGTGAT-3'; *aak-2*: 5'-AAGTCTGGAGTTGGGAATACG-3' and 5'-GTATGCATTCTTTGTGG AACC-3'; *hsf-1*: 5'-TCCGTATAAGAATGCGACTAGG-3' and 5'-TAGCTTCTG

ATGTGGTTGAAGG-3'; *hsp-1*: 5'-GGACGTCTTTCCAAGGATGA-3' and 5'-TCAAGATCTCGTCTGACTTG-3'; *hsp-16.2*: 5'-CTGTGAGACGTTGAGATTGATG-3' and 5'-CTTTACCACTATTTCCGTCAG-3'; *ash-2*: 5'-CGATCGAAACACGGAACGA-3' and 5'-TGCCGGAATCTGCAGTTTTT-3'; *set-2*: 5'-TCGAAGATTGAAGGTGAAGAGAG-3' and 5'-ATCATCTTTTTCGCGAACTGTAA-3'; *HSPD1*: 5'-TGCTGAGTTTGAATGAGCAA-3' and 5'-CAATCTGCTCTCAAATGGACA-3'; *Hsp90A1*: 5'-GAAATCTGTAGAACC CAAATTTCAA-3' and 5'-TCTTTGGATACCTAATGCGACA-3'; luciferase: 5'-ACGTCTTCCCGACGATGA-3' and 5'-GTCTTCCGTGCTCCAAAAC-3'.

RNA-sequencing analysis

Total RNA from four biological replicates of worms sorted at the L2 stage (extracted as described above) was assessed for quality using the TapeStation (Agilent). Samples were prepared using the Illumina TruSeq Stranded Total RNA Library Prep kit (Illumina). Total RNA (100 ng) was rRNA-depleted using Ribo-Gone (Takara Bio). The rRNA-depleted RNA was then fragmented and copied into first strand cDNA using reverse transcriptase and random primers. The products were purified and enriched by PCR (15 cycles) to create the final cDNA library. The 3' prime ends of the cDNA were adenylated and ligated to adapters, including a 6-nt barcode unique for each sample. Final libraries were checked for quality and quantity by TapeStation and qPCR using Kapa's library quantification kit for Illumina Sequencing platforms (Kapa Biosystems). The samples were pooled, clustered on an Illumina cBot and sequenced on one lane of an Illumina HiSeq4000 flow cell, as paired-end 50-nt reads. The quality of the raw reads data for each sample (for example, low-quality scores, over-represented sequences, inappropriate GC content) was checked using FastQC (v.0.11.3). The Tuxedo Suite software package was used for the computational analysis of the RNA sequencing^{29,30}. In brief, reads were aligned to the reference genome WS220 using TopHat (v.2.0.13) and Bowtie2 (v.2.2.1.). Cufflinks/CuffDiff (v.2.1.1) was used for expression quantitation, normalization, and differential expression analysis, using reference genome WS220. For this analysis, we used parameter settings: "–multi-read-correct" to adjust expression calculations for reads that map in more than one locus, as well as "–compatible-hits-norm" and "–upper-quartile-norm" for normalization of expression values. Diagnostic plots were generated using the CummeRbund R package. Genes and transcripts were identified as being differentially expressed based on three criteria: test status = "OK", FDR ≤ 0.05, and fold change ≥ ±1.5. The Bioconductor Package GSA was used to perform enrichment test analysis. The algorithm was modified from the original gene set enrichment analysis (GSEA)³¹, for better power. Gene sets were downloaded from sources indicated in Supplementary Table 1. All FDR corrected *P* values in this result are extremely significant (FDR ≈ 0).

Western blot

Standard methods for western blotting were used for the detection of proteins from worm lysates. In brief, 3,000–5,000 L2-staged worms were collected in 20 µl of M9 buffer and snap frozen in liquid nitrogen. Laemmli loading buffer was added to the worm pellet (1:1 volume) and the samples were boiled for 5 min, separated by SDS–PAGE and transferred to PVDF membranes. Blots were blocked for 1 h with 5% milk in PBS and probed with anti-H3 (Abcam, ab1791; 1:2,000), anti-H3K4me3 (Abcam, ab8580; 1:1,000), anti-H3K27ac (Abcam, ab4729; 1:1,000), anti-H3K27me3 (Millipore, 07-449; 1:1,000), anti-ASH-2 (Abmart, X3-GSEFZ3, 1:1,000;) or anti-β-tubulin (Santa Cruz, sc-5274; 1:2,000) primary antibodies overnight at 4 °C. For the extraction of mammalian proteins, HeLa cells were treated with trypsin (Invitrogen, 25200056) washed twice with PBS and collected in lysis buffer (RIPA buffer, 1 mM PMSF, protease inhibitor cocktail and 1 mM EDTA; 0.5 ml lysis buffer per 5 × 10⁶ cells). Samples were incubated for 45 min at 4 °C with constant agitation. Lysates were spun down (4 °C, 20 min, 12,000 rpm) and snap frozen in liquid nitrogen. Laemmli loading buffer was added to the lysates (1:1 volume) and the samples were boiled for

5 min, separated by SDS–PAGE and transferred to PVDF membranes. Blots were blocked for 1 h with 5% milk in PBS and probed with anti-H3 (Abcam, ab1791; 1:2,000), anti-H3K4me3 (Abcam, ab8580; 1:1,000), anti-ASH2L (Bethyl laboratories, polyclonal, A300-489A; 1:1,000), anti-MLL1 (Bethyl laboratories, polyclonal, A300-374A; 1:500) or anti-β-tubulin (Santa Cruz, sc-5274; 1:2,500) primary antibodies overnight at 4 °C. HRP conjugated anti-rabbit (ThermoScientific, 31460) and anti-mouse (ThermoScientific, 31430) secondary antibodies were used at 1:5,000 dilution for 1 h at room temperature. Proteins were detected using Clarity ECL western blotting substrate (BioRad) and signal was captured using a BioRad ChemiDoc Touch imaging system.

C. elegans RNAi

Escherichia coli HT115 (DE3) strains transformed with vectors expressing dsRNA of the genes of interest (*ash-2*, *set-2*, *wdr-5.1*, empty pL4440) were obtained from the Ahlinger library (a gift from G. Sankovszki), sequence-verified and grown at 37 °C as per manufacturer's recommendations. L1 worms obtained from synchronized populations were placed onto NGM plates containing ampicillin (100 mg ml⁻¹) and IPTG (0.4 mM) seeded with the respective bacteria. Worms were cultivated on either RNAi or the empty vector control bacteria for two generations.

Mammalian cell culture and H₂O₂ treatment

HeLa (EM-2-11ht) cells (a gift from J. Nandakumar and authenticated by STR) were cultured in DMEM (Life Technologies, 11995-065), supplemented with 10% Fetal Bovine Serum (Sigma-Aldrich, F4135) and 1% penicillin–streptomycin (Gibco, 15140-122) at 5% CO₂. At 80% confluency, cells were washed with PBS (Life Technologies, 10010023) and treated with HBSS (Life Technologies, 14025-092) supplemented with 0.1 mM or 0.3 mM H₂O₂ and incubated at 37 °C for 30 min.

Mammalian siRNA and heat-shock treatment

Eight thousand cells were transfected with 4.8 pmol of *ASH2L* siRNA (Dharmacon, M-019831-01-0005) or non-targeting siRNA (Dharmacon, D-001210-02-05) using Lipofectamine RNAiMax (Invitrogen, 13778-150) in OPTI-MEM I Reduced serum medium (Gibco, 31985-062). For heat-shock treatment, cells were washed with PBS 72 h after siRNA transfection and placed in HBSS. Plates were wrapped with parafilm and submerged in a pre-heated water bath at 43 °C for the indicated time points. Viability was determined using the CellTiter-Glo Kit (Promega) as per manufacturer's recommendations. Luminescence was monitored on a FLUOstar Omega microplate reader (BMG Labtech).

Histone methyltransferase activity assays

SET domains of SET1/MLL family proteins (MLL1, SET1A, SET1B), RBBP5 (full length), ASH2L (full length), and WDR5 (full length) were purified as previously described³². The purified proteins were diluted to 10 µM and incubated with 1 mM or 2 mM H₂O₂ at 4 °C for one hour in the buffer 25 mM Tris-HCl, pH 8.0. For DTT recovery, 4 mM DTT was added after H₂O₂ treatment and was incubated at 4 °C for 60 min. After oxidation, excess H₂O₂ was removed by ultrafiltration. Methyltransferase assays were performed using H3 peptides (residues 1–20) with one additional Tyr-residue at the C terminus for accurate quantification of peptides. An enzyme-coupled continuous spectrophotometric assay system was employed to monitor the time course of the reaction^{33,34}. This assay system, which monitors the appearance of the cofactor product (SAH) at an absorbance of 515 nm (that is, OD₅₁₅) contained the following components: 25 mM Tris (pH 8.0), 320 nM AdoHcy nucleosidase, 480 nM adenine deaminase, 40 U l⁻¹ xanthine oxidase, 20,000 U l⁻¹ horseradish peroxidase, 4.5 mM 3,5-dichloro-2-hydroxybenzene-sulfonic acid, 0.894 mM 4-aminophenazone, 40 µM MnCl₂, 2.25 µM K₄Fe(CN)₆·3H₂O, 200 µM S-adenosyl-methionine and 1 µM of the four mammalian proteins that constitute the minimal H3K4-methylating complex. All components were mixed in 30 µl volume in 384-well plate at room temperature, and the reaction was initiated by adding 400 µM

Article

H3 peptide substrate. The OD₅₁₅ was monitored using a Synergy Neo Multi-Mode Reader (Bio-Tek) for 1 h at 28 °C. The slope of OD₅₁₅ versus time from the first 20 min linear range was converted into reaction rates. The relative activity for each complex without any H₂O₂ pretreatment was set to 1. A buffer control was used to determine the baseline.

In vitro protein oxidation and thiol trapping

Purified MLL1 SET domain (15 µM) was treated with either 2 mM DTT or 2 mM H₂O₂ for 30 min at 4 °C or 30 °C. To stop the reaction, the H₂O₂-treated samples were mixed with catalase (0.5 mg ml⁻¹). To reduce reversible thiol modifications the oxidized sample were treated with 4 mM DTT for 30 min at 30 °C. The reduced cysteines were blocked with 20 mM NEM (*N*-ethylmaleimide) before SDS–PAGE analysis. For reverse thiol trapping experiments, the samples were resuspended in a denaturing thiol-trapping buffer (2.3 M urea, 0.2% SDS, 10 mM EDTA, 200 mM Tris-HCl, pH 8.5) supplemented with 20 mM NEM for 30 min at 25 °C. Proteins were precipitated with 10% trichloroacetic acid (TCA). After centrifugation, the pellets were washed with 10% TCA and 5% TCA and re-dissolved in the denaturing thiol-trapping buffer supplemented with 4 mM DTT to reduce reversible thiol modifications. After 45 min of incubation at 30 °C, all new cysteine thiols were labelled with 25 mM AMS for 5 min at 25 °C. Proteins were analysed on SDS–PAGE under non-reducing conditions and visualized using silver staining. For the mass spectrometry analysis of the cysteine-containing peptides, iodoacetamide (IAM) was used instead of AMS to label reversibly oxidized cysteines. After SDS–PAGE under non-reducing conditions and Coomassie staining, protein bands were cut out, trypsin-digested, and analysed by nano liquid chromatography with tandem mass spectrometry (LC–MS/MS; MS Bioworks).

Statistical analysis

The Prism software package (GraphPad Software 7) and the Microsoft Office 2010 Excel software package (Microsoft Corporation) were used to carry out statistical analyses. Information about statistical tests, *P* values and *n* numbers are provided in the respective figures and figure legends.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All relevant data are available and/or included with the manuscript as Source Data or Supplementary Information. RNA-sequencing data

have been uploaded to the Gene Expression Omnibus (GEO) database with accession number GSE138502.

26. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
27. Stroustrup, N. et al. The *Caenorhabditis elegans* Lifespan Machine. *Nat. Methods* **10**, 665–670 (2013).
28. Koopman, M. et al. A screening-based platform for the assessment of cellular respiration in *Caenorhabditis elegans*. *Nat. Protocols* **11**, 1798–1816 (2016).
29. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
30. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
31. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
32. Ng, S. B. et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010).
33. Dorgan, K. M. et al. An enzyme-coupled continuous spectrophotometric assay for S-adenosylmethionine-dependent methyltransferases. *Anal. Biochem.* **350**, 249–255 (2006).
34. Southall, S. M., Cronin, N. B. & Wilson, J. R. A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. *Nucleic Acids Res.* **42**, 661–671 (2014).

Acknowledgements We thank M. Malinouski and T. Mullins for assistance with the reconfiguration of the Biosorter; G. Csankovszki for antibodies, *C. elegans* RNAi feeding clones and comments; B. Braeckman for the N2jrls2[Prpl-17::Grx-1-roGFP2] strain; J. Nandakumar for HeLa (EM-2-11ht) cells; the *Caenorhabditis* Genetics Center (funded by National Institutes of Health Infrastructure Program P40 OD010440) for *C. elegans* strains; the DNA Sequencing Core (BRCF), R. Tagett, W. Wu and the Bioinformatics Core of University of Michigan for RNA sequencing and data analysis; K. Wan for protein purification; R. Sawarkar and J. Labbadia for important suggestions; Jakob laboratory members for comments on the manuscript and J. Bardwell for critically reading the manuscript. Mass spectrometry was performed by MS Bioworks. This work was supported by NIH grants GM122506 and AG046799 as well as the Priority Program SPP 1710 of the Deutsche Forschungsgemeinschaft (Schw823/3-2) to U.J., a NIH T32 Career Training in the Biology of Aging grant to D.B., a NIH T32 Career Training in the Biology of Aging grant and a Bright Focus ADR Fellowship (A2019250F) to B.J.O., and the National Natural Science Foundation of China (31470737) to Y.C.

Author contributions D.B. conceived and conducted most experiments, performed data analysis and wrote the manuscript; D.K. conceived experiments and initiated work with the BioSorter; Y.Z. performed the in vitro methyltransferase assays; K.U. performed the reverse thiol trapping and prepared samples for mass spectrometry analysis; B.J.O. assisted with RNAi experiments and western blots for methylation marks in *C. elegans*; L.X. performed siRNA in HeLa cells; M.K. built and operated the lifespan instrument; A.K. assisted with worm sorting and produced brood size data; Y.-T.L. purified mammalian proteins; Y.D. conceived experiments and provided material; S.Q. and Y.C. conceived experiments; U.J. conceived experiments, conducted data analysis and wrote the manuscript.

Competing interests The authors declare no competing interests.

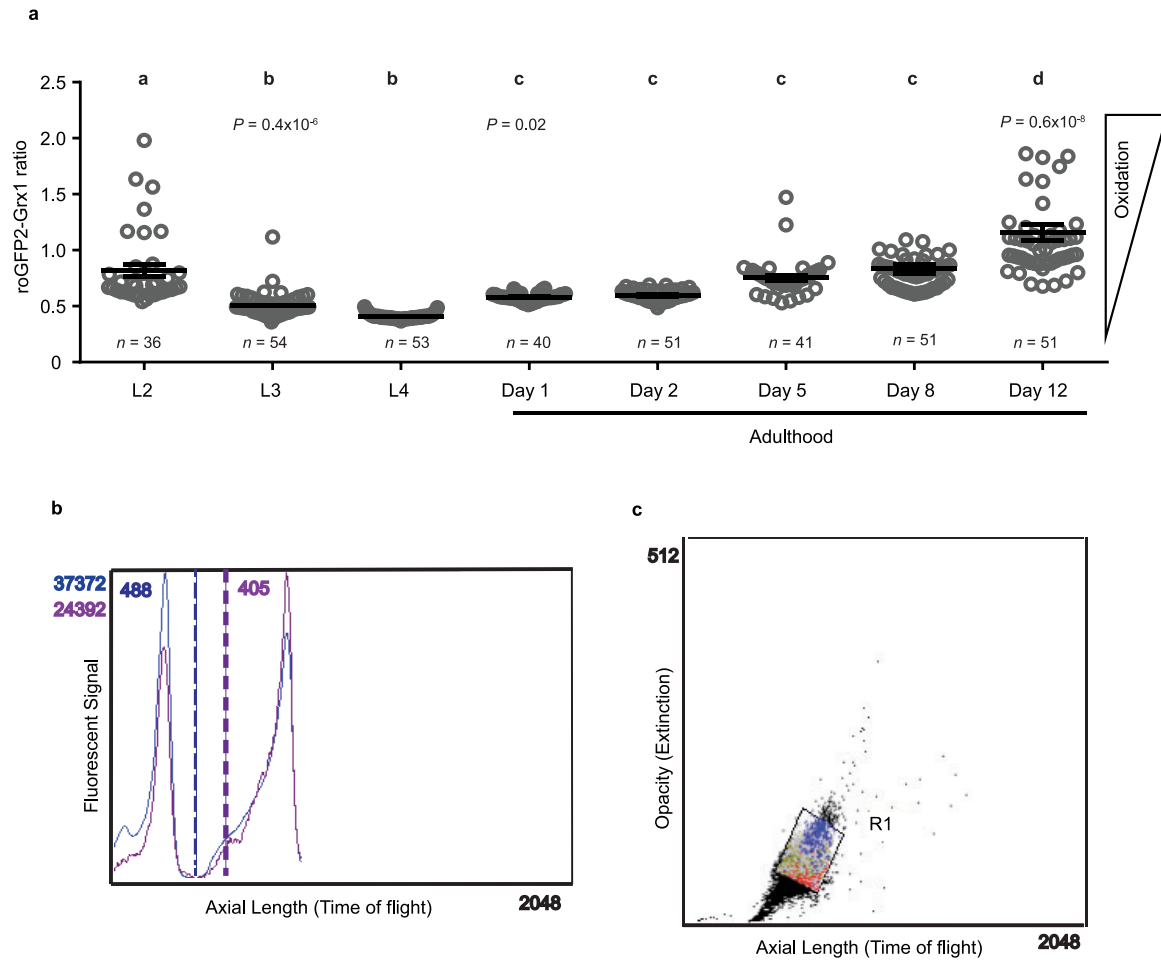
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1814-y>.

Correspondence and requests for materials should be addressed to U.J.

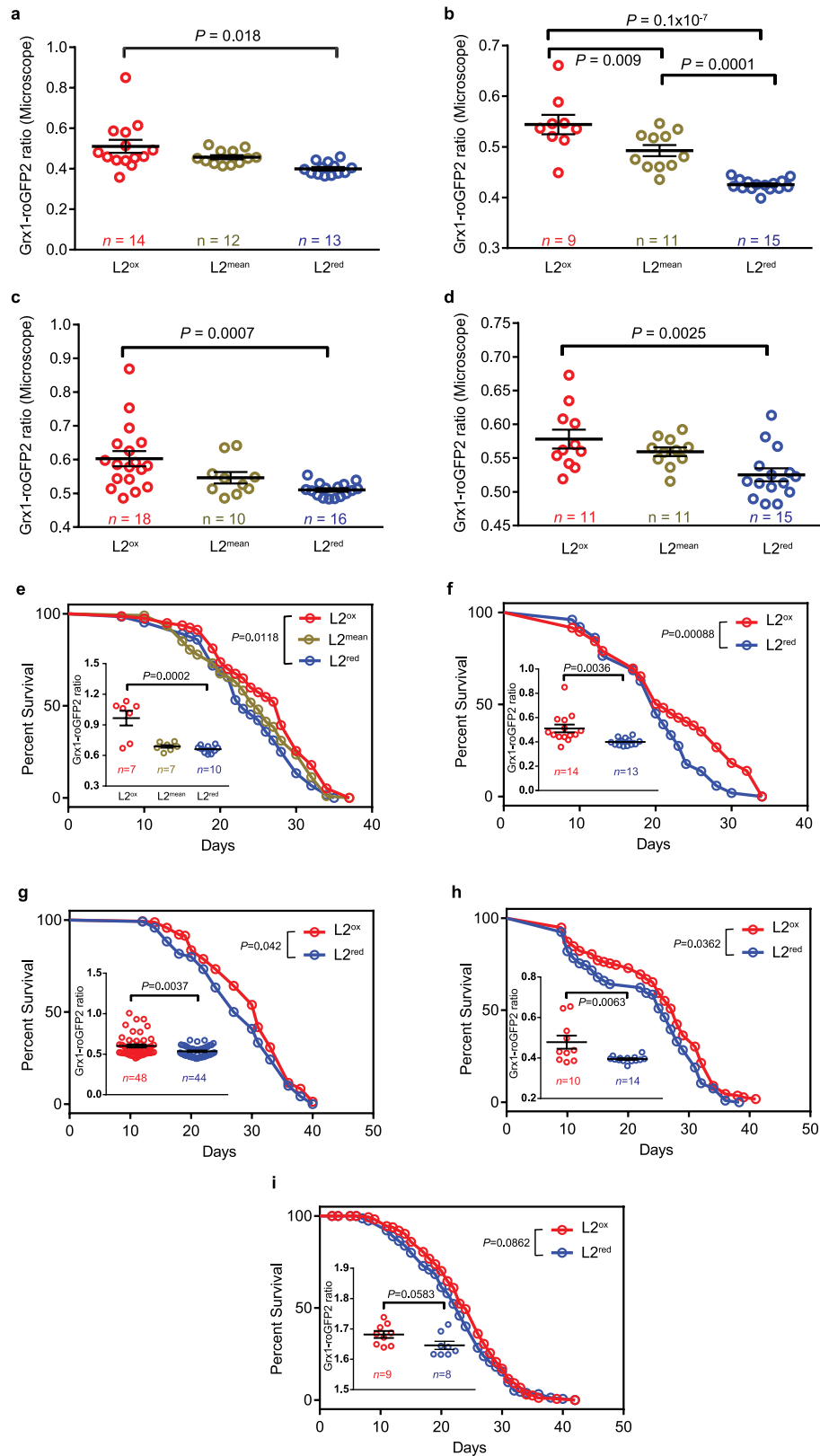
Peer review information Nature thanks Anne Brunet, Michael Ristow and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



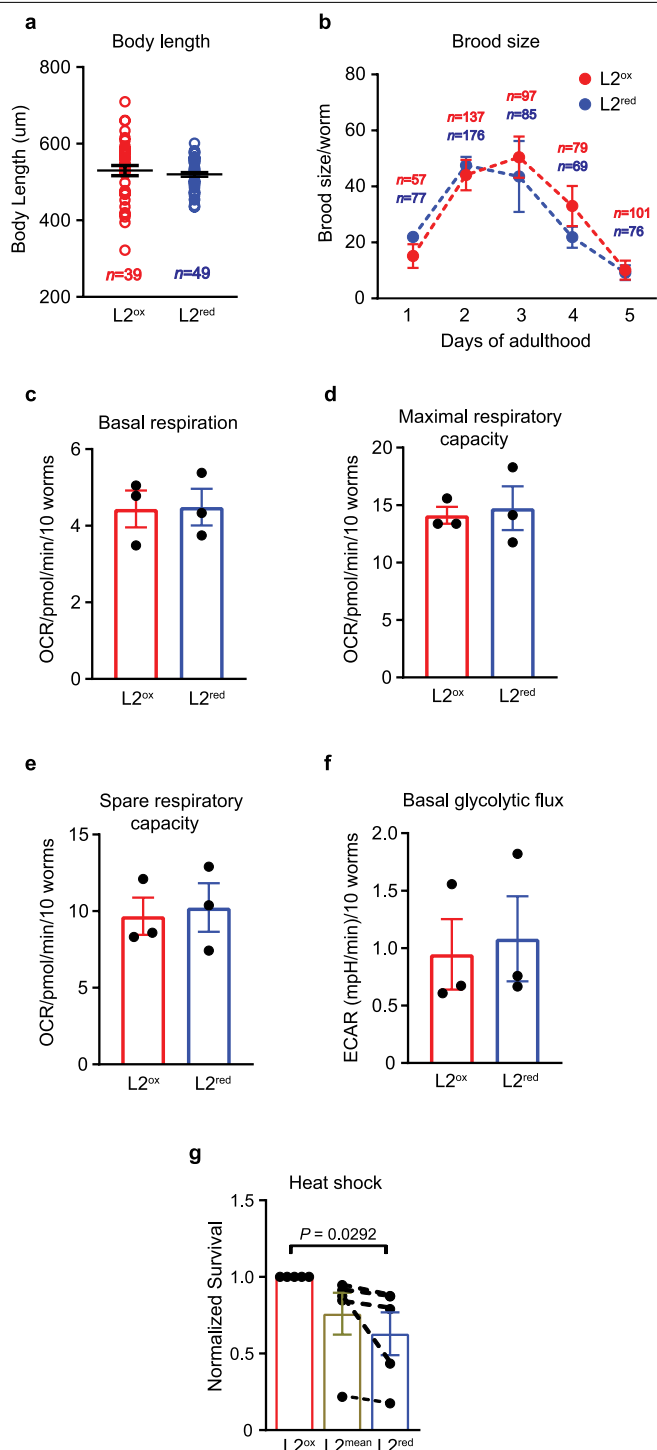
Extended Data Fig. 1 | In vivo readout of endogenous redox states at different stages during *C. elegans* lifespan and sorting parameters of oxidized and reduced subpopulations. **a, Microscopy analysis of the Grx1-roGFP2 ratio of individual *N2jrls2[Prpl-17::Grx1-roGFP2]* worms (symbol) cultivated at 15 °C and imaged at the indicated time points. Data points that are not significantly different from each other ($P > 0.05$) are labelled with the same**

letter. Data are mean \pm s.e.m; n , number of worms; one-way ANOVA with Tukey correction. **b**, The Grx-roGFP2 ratio (A_{405}/A_{488}) was calculated using the partial profiling feature (pp) configured to analyse extinction and emission data from 488 nm and 405 nm lasers that sequentially excited each worm. **c**, A population of *N2jrls2[Prpl-17::Grx1-roGFP2]* at the L2 stage, separated based on opacity (extinction) and length (time of flight) was gated as R1.

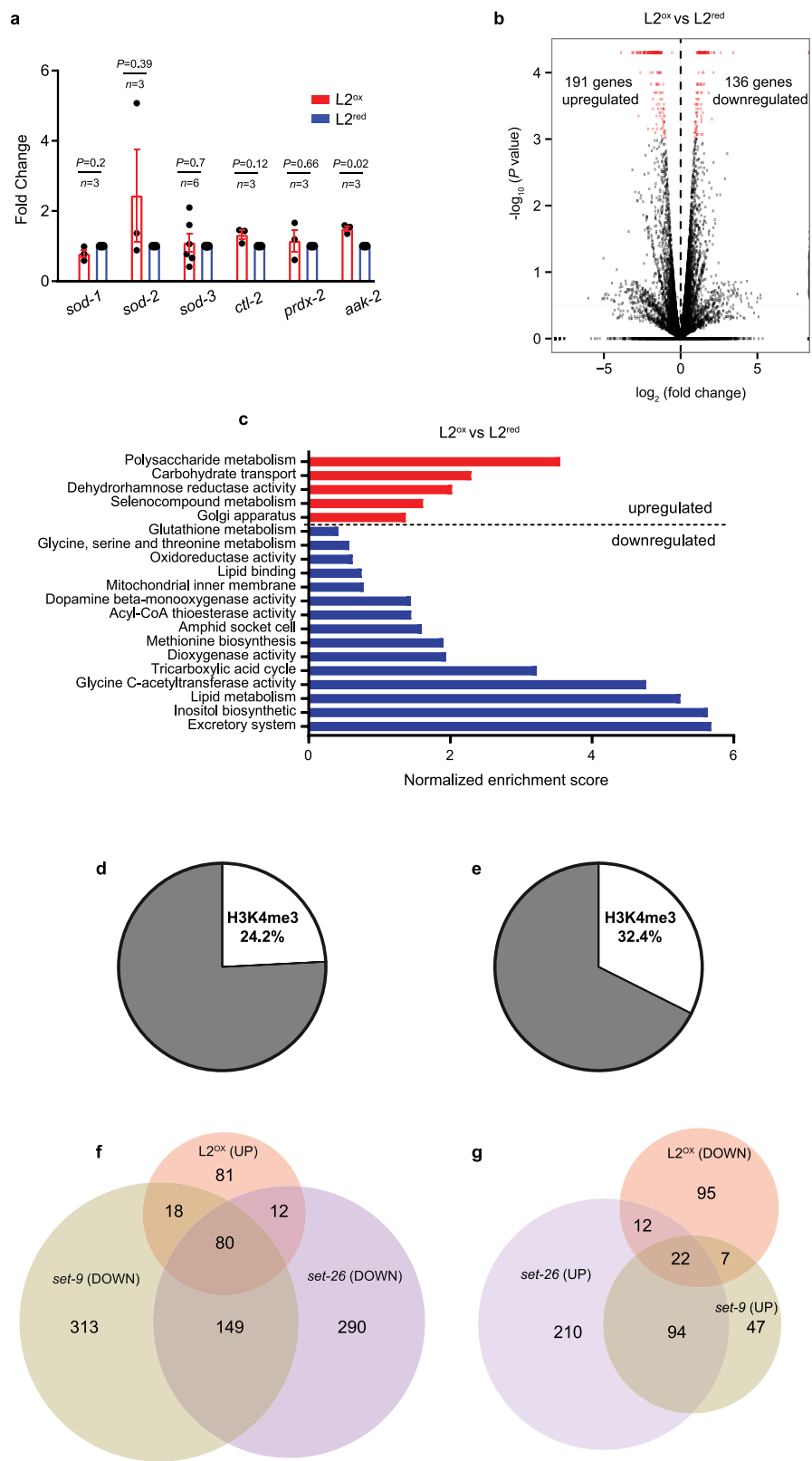


Extended Data Fig. 2 | Sorting efficiency and lifespan of $L2^{ox}$ and $L2^{red}$ subpopulations. **a–d**, Microscopy analysis of the Grx1-roGFP2 ratio of individual worms (dots) previously sorted into $L2^{ox}$, $L2^{mean}$ and $L2^{red}$ subpopulations. n , number of worms; one-way ANOVA with Tukey correction. **e–i**, Survival curves of sorted $L2^{ox}$, $L2^{mean}$ and $L2^{red}$ worms. For n numbers,

P values (log-rank test) see Extended Data Table 2. Insets, Grx1-roGFP2 ratio of individual worms (dots), assessed by fluorescence microscopy after sorting. n , number of worms. For P values (two-sided unpaired t -test) see Extended Data Table 2.



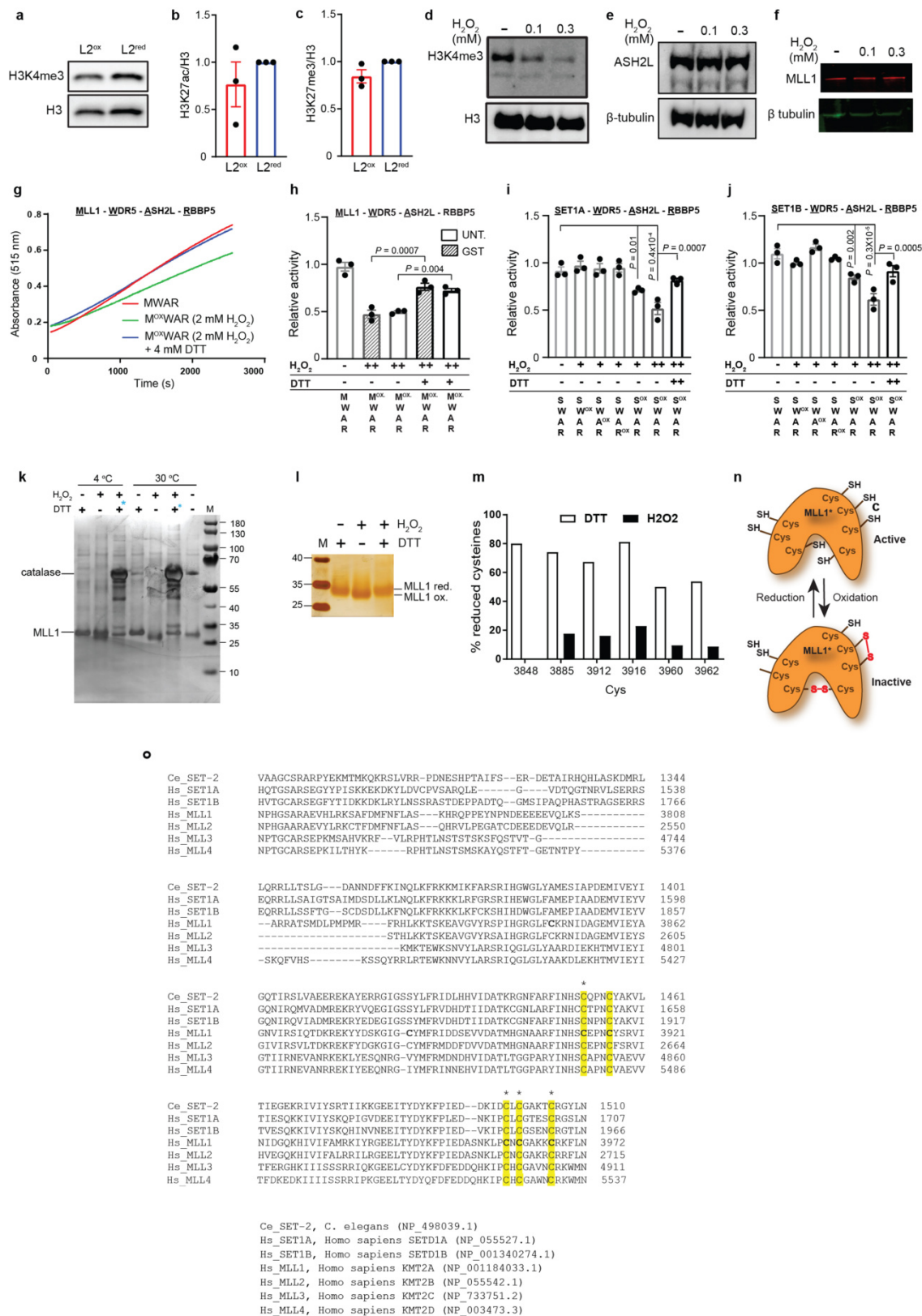
Extended Data Fig. 3 | Physiological properties of L2^{ox} and L2^{red} sorted worms. **a**, Length measurements of L2^{ox} and L2^{red} worms (symbol) from nose to tail tip immediately after sorting. No significant difference; $P = 0.4735$ (unpaired two-sided t -test). **b**, Brood size of L2^{ox} and L2^{red} worms, measured at the indicated time points. n , number of worms. No significant difference within a single age; $P = 0.6532$ (two-way ANOVA). **c–f**, Basal respiration (**c**), maximal (**d**) and spare (**e**) respiratory capacity and basal rates of flux through glycolysis (**f**) of L2^{ox} and L2^{red} worms. $n = 3$ independent sorting experiments. ECAR, extracellular acidification rate; OCR, oxygen consumption rate. $P = 0.9469$ (**c**), $P = 0.7784$ (**d**), $P = 0.7904$ (**e**) and $P = 0.7925$ (**f**); two-sided unpaired t -test. **g**, Survival of L2^{ox}, L2^{mean} and L2^{red} worms 20 h after heat shock. $n = 5$ independent sorting experiments; two-sided unpaired t -test. The connected data points represent data from the same sorting experiment. The survival of L2^{ox} is set to 1. All data are mean \pm s.e.m.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Gene expression profiles of L2^{ox} and L2^{red}. **a**, Steady-state transcript levels of selected oxidative stress-related genes in L2^{ox} and L2^{red} worms. *n*, number of independent sorting experiments; unpaired two-sided *t*-test. Data are mean ± s.e.m. **b**, Volcano plot showing fold changes versus *P* values for the transcriptomes of L2^{ox} and L2^{red} subpopulations. DEGs (*P* ≤ 0.05) are represented by red dots (see Methods for statistical definition of DEGs). Data were collected from four independent sorting experiments. **c**, GSEA of the 327 DEGs. Normalized enrichment scores (see Methods for calculation) are represented by the bar graph. Terms (for summary, see Supplementary Table 1) indicating origin, process or phenotype associated with genes known to have a role in the process are shown on the left. Some terms (*) have been merged and are represented as a single category bar for

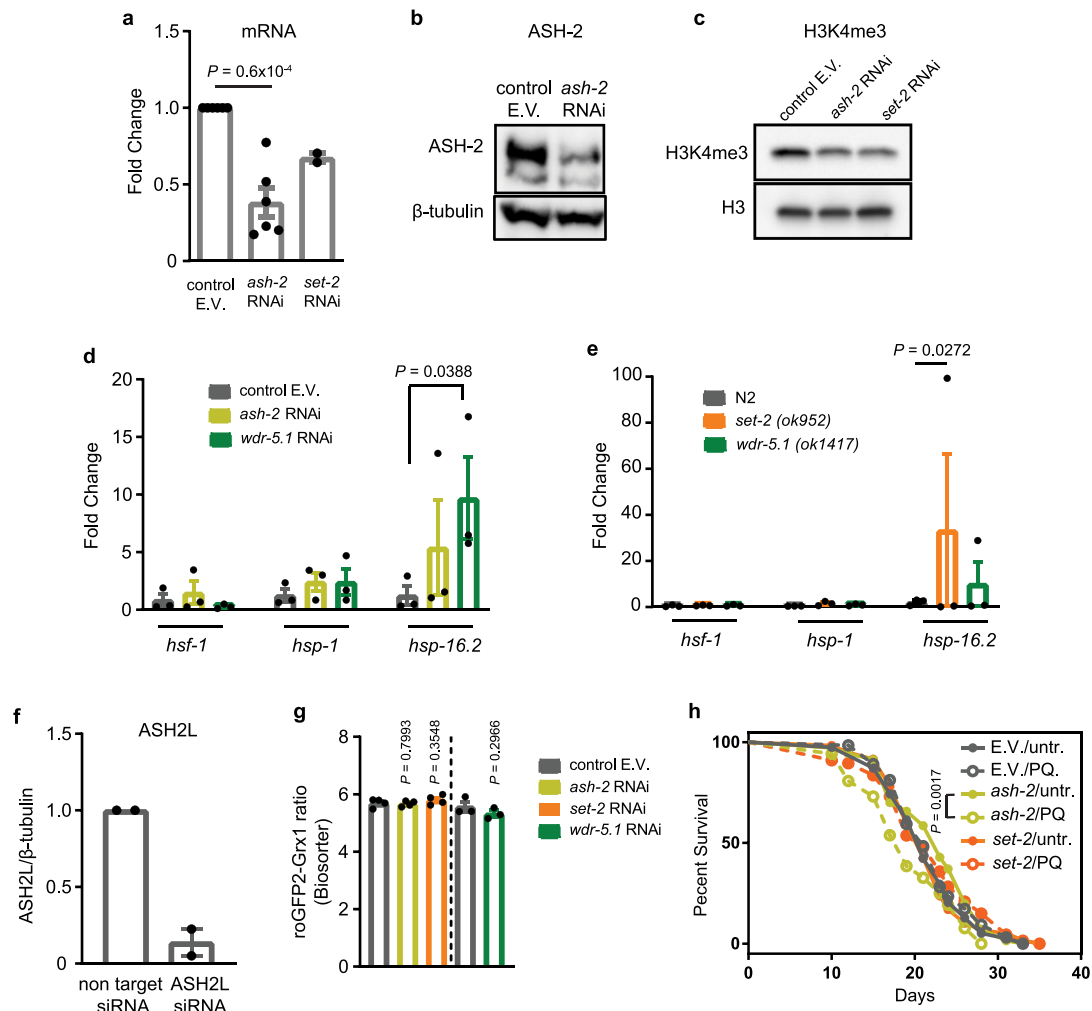
simplicity (for detailed values, see Supplementary Table 1). **d**, **e**, Percentage of DEGs identified in L2^{ox} that intersect with H3K4me3 peak signals within their 5' region (500 bp upstream and downstream from the transcription start site). The H3K4me3 chromatin immunoprecipitation (ChIP) datasets were generated from L3-staged N2 worms: ChIP chip, GEO entry GSE30789 in **d** and chromatin immunoprecipitation followed by sequencing (ChIP-seq), GEO entry GSE28770 in **e**, indicating that these marks are set during larval development. Hypergeometric probability: **d**, *P* = 0.064; **e**, *P* = 2.786 × 10⁻⁶. In **f**, **g**, Venn diagrams show the overlap among upregulated (**f**) or downregulated (**g**) gene sets in L2^{ox} and downregulated or upregulated *set-9(rw5)* and *set-26(tm2467)* gene sets (GEO entry: GSE100623). See Supplementary Table 1 for datasets in **d–g**.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Redox sensitivity of in vivo H3K4me3 levels and in vitro histone methyltransferase complex activity. **a**, Global H3K4me3 levels in the sorted L2^{ox} and L2^{red} worms. A representative western blot using antibodies against H3K4me3 is shown. **b, c**, Quantification of global H3K27ac (**b**) and H3K27me3 (**c**) levels by western blot. $n = 3$ independent sorting experiments. $P = 0.3793$ (**b**) and $P = 0.0905$ (**c**); unpaired two-sided *t*-test. Data represent mean \pm s.e.m. **d–f**, Global H3K4me3 (**d**), ASH2L (**e**) and MLL1 (**f**) levels in HeLa cells before and after H₂O₂ treatment, as assessed by western blot. **g**, Time course of the in vitro methyltransferase reaction for core COMPASS subunits (SET domain of MLL1–WDR5–ASH2L–RBBP5). Reaction rates were derived from the first 20 min of the linear range. **h–j**, In vitro histone methyltransferase assays of core COMPASS subunits, consisting of purified GST–WDR5 (WDR5), GST–ASH2L (ASH2L), GST–RBBP5 (RBBP5) and either GST–MLL1SET domain or untagged MLL1SET domain (**h**), GST–SET1A SET domain (**i**) or GST–SET1B SET domain (**j**). Superscript ox indicates that the protein was pre-treated with either 1 mM (+) or 2 mM (++) H₂O₂ for 30 min before the activity assay. DTT was added after the H₂O₂ treatment. $n = 3$ independent experiments; one-way ANOVA with Sidak correction. Data are mean \pm s.e.m. **k**, The MLL1SET

domain was treated with either 2 mM DTT, 2 mM H₂O₂ or 2 mM H₂O₂, followed by 4 mM DTT. Catalase was used to quench the H₂O₂. The proteins were denatured and thiols were modified with NEM before loading onto non-reducing SDS–PAGE to prevent non-specific thiol oxidation. The proteins were visualized by silver staining. M, marker. **l**, MLL1SET domain treated with either 2 mM DTT, 2 mM H₂O₂ or 2 mM H₂O₂, followed by 4 mM DTT. All reduced protein thiols were then labelled with the 500-Da thiol-reactive compound AMS, causing a 500-Da mass decrease per oxidized thiol—detectable on reducing SDS–PAGE. M, protein marker. **m**, Cysteine oxidation state in MLL1SET domain after treatment with either 2 mM DTT or 2 mM H₂O₂ followed by NEM labelling as assessed by LC–MS/MS. The peptide containing Cys3967 could not be detected. **n**, Schematic representation of the redox sensitivity of the MLL1SET domain. For blot and gel source images, see Supplementary Figs. 1 and 3. **o**, Sequence alignment of the SET domain. All cysteines in MLL1 are shown in bold, and the five absolutely conserved cysteines are highlighted in yellow. Cysteines shown to be involved in zinc coordination are marked with an asterisk. NCBI protein BLAST and Clustal Omega Multiple Sequence Alignment, Clustal O (1.2.4) were used.



Extended Data Fig. 6 | Effects of H3K4me3 downregulation on heat-shock response and endogenous redox state. a, ASH-2 and SET-2 transcript levels of *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with *ash-2* or *set-2* RNAi for 2 generations. $n = 6$ (*ash-2*) and $n = 2$ (*set-2*) independent experiments; unpaired two-sided *t*-test. E.V., empty vector. **b**, ASH-2 protein levels in *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with control RNAi or *ash-2* RNAi for two generations using western blot analysis. **c**, H3K4me3 levels in *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with control RNAi, *ash-2* RNAi or *set-2* RNAi for two generations. **d**, Transcript levels of selected heat-shock genes after heat-shock treatment of *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with the indicated RNAi. $n = 3$ independent experiments; one-way ANOVA with Bonferroni correction. **e**, Transcript levels of selected heat-shock

genes in *set-2* or *wdr-5.1* mutants before and after heat-shock treatment. $n = 3$ independent experiments; one-way ANOVA with Bonferroni correction. **f**, ASH2L levels following ASH2L siRNA treatment of HeLa cells. $n = 2$ independent experiments. **g**, Grx1-roGFP2 ratios of L2 larval worms treated with *ash-2* RNAi, *set-2* RNAi, *wdr-5.1* RNAi or the empty vector were measured using the Biosorter. $n = 4$ (*ash-2*, *set-2*) and $n = 3$ (*wdr-5.1*) independent sorting experiments; unpaired two-sided *t*-test. **h**, Representative survival curves of *N2jrls2[Prpl-17::Grx1-roGFP2]* worms treated with *ash-2* or *set-2* RNAi for two generations and treated with 1 mM paraquat (PQ) for 10 h at the L2 larval stage. For n numbers, repetitions and statistics (log-rank), see Extended Data Table 4. Data in **a**, **d**–**g**, represent mean \pm s.e.m. For blot source images, see Supplementary Figs. 1, 3.

Extended Data Table 1 | Lifespan assays of L2^{ox}, L2^{mean} and L2^{red} subpopulations following heat-shock treatment or in the continuous presence of paraquat or juglone

Condition	Experiment	Sorted population	Mean	Max	Dead/ censored worms	P value	% change	Figure on text	Ratio BioSorter [‡]	P value (BioS.)
Heat shock (long term)	1	L2 ^{ox}	19	32.5	24				7.44±0.01(n=238)	
		L2 ^{mean}	17	31.5	19/4	0.639	10.5%*	Figure 2a	7.37±0.02(n=135)	0.002
		L2 ^{red}	13	27.3	32/1	0.0088	31.6%*		7.29±0.02(n=132)	0.2x10 ⁻⁸
	2	L2 ^{ox}	17.5	25	29/13				7.06±0.01(n=698)	
		L2 ^{red}	12	21	20/3	0.0077	31.4%*		7.01±0.012(n=319)	0.002
	3	L2 ^{ox}	18	25.9	96/2				7.15±0.007(n=893)	
		L2 ^{red}	15	24.2	77/1	0.0153	16.7%*		7.12±0.012(n=914)	0.009
Paraquat (2 mM)	1	L2 ^{ox}	17	26	13			Figure 2b	7.09±0.02(n=116)	
		L2 ^{mean}	14	26	11	0.4456	17.6%*		6.99±0.033(n=113)	0.006
		L2 ^{red}	10	21	20	0.0215	41.2%*		6.67±0.045(n=108)	0.2x10 ⁻¹³
	2	L2 ^{ox}	13	18.5	38				7.35±0.017(n=158)	
		L2 ^{red}	11	18.5	40	0.0456	15.4%*		7.14±0.011(n=277)	0.1x10 ⁻¹⁹
	3	L2 ^{ox}	13	18.6	47				6.71±0.02(n=139)	
		L2 ^{red}	12	15.5	38	0.0037	16.7% [†]		6.33±0.06(n=100)	0.8x10 ⁻⁷
Juglone (250 µM)	1	L2 ^{ox}	23	32.4	72				6.59±0.012(n=428)	
		L2 ^{red}	19	32.4	73	0.0033	17.4%*		6.55±0.012(n=513)	0.007
	2	L2 ^{ox}	18	33.3	86				7.43±0.01(n=186)	
		L2 ^{red}	18	29.5	59	0.0202	11.4% [†]		7.27±0.015(n=161)	0.1x10 ⁻¹⁶

Worms were sorted onto NGM plates and exposed to heat-shock treatment or the indicated compounds. *P* values for Kaplan–Meier survival analysis were calculated on the basis of the log-rank (Mantel–Cox) method. Samples were compared with L2^{ox} subpopulations. Maximum lifespan was defined as the lifespan of the last 10% of the population.

*Percentage change based on mean lifespan.

[†]Percentage change based on maximum lifespan.

[‡]Grx1-roGFP2 ratio of L2^{ox} and L2^{red} subpopulations, based on BioSorter analysis during sorting. Data are mean ± s.e.m. *n*, number of worms. Two-sided unpaired *t*-test.

Extended Data Table 2 | Lifespan assays of L2^{ox}, L2^{mean} and L2^{red} subpopulations

Exp.	Sorted population	Mean	Max	Dead/ censored worms	P value (survival)	% change	Figure on text	Ratio BioSorter [§]	P value (BioS.)	Ratio Microscope	P value (Micr.)
1	L2 ^{ox}	30	35.8	97	0.007	13.3%*	Fig. 2c	6.88±0.02(n=134)	0.1×10 ⁻⁸	0.49±0.008(n=36)	0.1×10 ⁻⁶
	L2 ^{mean}	26	33.5	74				6.69±0.02(n=143)		0.45±0.003(n=49)	
	L2 ^{red}	25	32.3	58				6.62±0.04(n=131)		0.44±0.003(n=50)	
2	L2 ^{ox}	28	35.1	78/2	0.1018	10.7%*		7.14±0.02(n=127)	0.532	0.96±0.07(n=7)	0.0027
	L2 ^{mean}	25	34.3	105/3				7.11±0.03(n=133)		0.69±0.01(n=7)	
	L2 ^{red}	23	33.5	63/1				6.94±0.07(n=136)		0.66±0.01(n=10)	
3	L2 ^{ox}	21	34	94/2	0.0088	12% [†]		7.45±0.02(n=139)	0.0003	0.51±0.03(n=14)	0.0036
	L2 ^{red}	20	30	51				7.35±0.01(n=716)		0.45±0.01(n=13)	
4	L2 ^{ox}	31	39.4	162/3	0.042	8%*		6.61±0.02(n=170)	0.0456	0.6±0.02(n=48)	0.0037
	L2 ^{red}	28.5	38.8	120				6.55±0.02(n=161)		0.54±0.01(n=44)	
5	L2 ^{ox}	28	36	114/4	0.0362	7%*		7.29±0.01(n=191)	0.3×10 ⁻⁹	0.48±0.03(n=10)	0.0063
	L2 ^{red}	26	35.4	106				7.12±0.02(n=88)		0.39±0.004(n=14)	
6 [¶]	L2 ^{ox}	24	34	164/7	0.0862 [‡]	4%*		6.19±0.02(n=275)	0.5×10 ⁻⁷	1.68±0.01(n=9)	0.0583
	L2 ^{red}	23	34	155/9				6.03±0.02(n=273)		1.65±0.01(n=8)	
7	E.V. ^{ox}	24	34	46	0.0004	12.5%*	Fig. 4d	5.55±0.04(n=114)	0.5×10 ⁻²⁸		
	E.V. ^{red}	21	28.3	63				4.49±0.067(n=114)			
8	E.V. ^{ox}	24	33	77/3	0.002	17%*	Fig. 4e	6.37±0.035(n=60)	0.1×10 ⁻⁷		
	E.V. ^{red}	20	31	80				5.85±0.076(n=63)			
9	E.V. ^{ox}	20	33	85	0.0264	15%*		6.08±0.046(n=133)	0.0032		
	E.V. ^{red}	17	30.7	77				5.87±0.052(n=107)			
10	E.V. ^{ox}	18	28.7	89/1	0.0077	11.5% [†]		6.03±0.022(n=225)	0.3×10 ⁻⁶²		
	E.V. ^{red}	18	25.4	85				5.10±0.038(n=226)			

Worms were sorted onto NGM plates and assessed for lifespan. *P* values for Kaplan–Meier survival analysis were calculated on the basis of the log-rank (Mantel–Cox) method unless noted otherwise. Samples were compared with L2^{ox} subpopulations. Maximum lifespan was defined as the lifespan of the last 10% of the population. E.V., empty vector (control RNAi conditions for experiments in Extended Data Table 4).

*Percentage change based on mean lifespan.

[†]Percentage change based on maximum lifespan.

[‡]*P* value for survival analysis calculated based on the Gehan–Breslow–Wilcoxon method.

[§]Grx1–roGFP2 ratio of L2^{ox} and L2^{red} subpopulations, based on BioSorter analysis during sorting. Data represent mean ± s.e.m. *n*, number of worms. Two-sided unpaired *t*-test.

^{||}Grx1–roGFP2 ratio of L2^{ox} and L2^{red} subpopulations based on microscopy analysis, following sorting. Data are mean ± s.e.m. *n*, number of worms. Two-sided unpaired *t*-test.

[¶]Lifespan assay conducted with experimenter blind to the type of worms assayed.

Extended Data Table 3 | Lifespan assays of L2^{ox} and L2^{red} subpopulations following transient exposure to oxidizing or reducing conditions

Experiment	Sorted population	Mean	Max	Dead/ censored worms	P value	% change	Figure on text	Ratio BioSorter [§]	P value (BioS.)
1	L2 ^{ox}	24	34.4	91				7.07±0.018(n=317)	
	1 mM PQ	26	33.7	85	0.4364	7.7%*	Figure 2f		
	10 mM NAC	19	32.9	99	0.008	20.8%*			
	L2 ^{red}	21	31.2	109				7.±0.011(n=402)	0.00074
	1 mM PQ	23	31.7	139/1	0.0261	8.7%*	Figure 2e		
	10 mM NAC	21	29.4	157	0.0546	-			
2	L2 ^{ox}	16	23.3	86				7.09±0.007(n=800)	
	L2 ^{red}	15	19	54	0.0002	18.4% [†]		7.05±0.009(n=1218)	0.00034
	L2 ^{red} + 1 mM PQ	15.5	25.6	64	0.0002 [‡]	9% [†]			
3	L2 ^{ox}	23	29.8	89					
	10 mM NAC	21	30.4	50	0.05 [‡]	8.7%*			
4 [¶]	L2 ^{Mixed}	15	21.1	66					
	0.1 mM PQ	15	20.6	107	0.5137	2.4% [†]	Figure 2d		
	1 mM PQ	15	27.8	96	0.0175	24.1% [†]			
	10 mM NAC	14.5	19.5	66	0.3667	7.6% [†]			
5 [¶]	L2 ^{Mixed}	12	16	84					
	0.1 mM PQ	12	20.7	101	0.003	22.7% [†]			
	1 mM PQ	12	29.7	88	0.0023	46.1% [†]			
	10 mM NAC	11	18.7	101	0.1809	14.4% [†]			

Worms were sorted into liquid medium, exposed to paraquat or NAC for 10 h and then returned to NGM plates for lifespan assessment. P values for Kaplan–Meier survival analysis were calculated on the basis of the log-rank (Mantel–Cox) method unless noted otherwise. Samples were compared with L2^{ox} subpopulations or L2^{Mixed} populations. Maximum lifespan was defined as the lifespan of the last 10% of the population.

*Percentage change based on mean lifespan.

[†]Percentage change based on maximum lifespan.

[‡]P value for survival analysis calculated based on the Gehan–Breslow–Wilcoxon method.

[§]Grx1-roGFP2 ratio of L2^{ox} and L2^{red} subpopulations at the L2 stage, based on BioSorter analysis during sorting. Data are mean ± s.e.m. n, number of worms. Two-sided unpaired t-test.

[¶]Sample compared with the reduced subpopulation.

[¶]Lifespan assays performed in the lifespan machine (see Methods).

Extended Data Table 4 | Lifespan assays upon H3K4me3-targeting RNAi treatment

Exp.	Sorted population	Mean	Max	Dead/ censored worms	P value	% change	Figure on text	Ratio BioSorter [‡]	P value (BioS.)
1 ^{&}	<i>ash-2</i> RNAi ^{ox}	24	33.3	40/2			Figure 4d	5.00±0.043(n=95)	
	<i>ash-2</i> RNAi ^{red}	25	34	45/4	0.1373	4.2%*		4.66±0.046(n=85)	0.1×10 ⁻⁶
2 ^{&}	<i>set-2</i> RNAi ^{ox}	22	32	57/1			Figure 4e	6.17±0.047(n=52)	
	<i>set-2</i> RNAi ^{red}	22	30	72/1	0.7739	-		5.64±0.03(n=224)	0.2×10 ⁻¹⁴
3 ^{&}	<i>ash-2</i> RNAi ^{ox}	30	39.1	70				6.12±0.03(n=186)	
	<i>ash-2</i> RNAi ^{red}	30	37.1	79	0.5721	-		5.8±0.039(n=216)	0.9×10 ⁻⁹
	<i>set-2</i> RNAi ^{ox}	20	34	76				6.13±0.019(n=259)	
	<i>set-2</i> RNAi ^{red}	20	33	55	0.7339	-		5.67±0.026(n=295)	0.1×10 ⁻³⁸
4 ^{&}	<i>ash-2</i> RNAi ^{ox}	25	30.2	87				6.17±0.021(n=223)	
	<i>ash-2</i> RNAi ^{red}	25	30.4	79	0.9133	-		5.56±0.046(n=223)	0.5×10 ⁻²⁶
	<i>set-2</i> RNAi ^{ox}	18	25.7	76				6.19±0.025(n=214)	
	<i>set-2</i> RNAi ^{red}	18	27	85	0.8589	4.8% [†]		5.29±0.041(n=217)	0.1×10 ⁻⁵²
5	control E.V. untr.	21	30	78			Ext. Data Fig. 6h		
	control E.V. PQ	21	31.7	64	0.5073	-			
	<i>ash-2</i> RNAi untr.	23	30.4	70					
	<i>ash-2</i> RNAi PQ	19	27.6	52	0.0017	17.4%*			
	<i>set-2</i> RNAi untr.	21	31.3	59					
	<i>set-2</i> RNAi PQ	21	32.1	67	0.4515	-			
6	control E.V. untr.	20	29.5	58					
	control E.V. PQ	18	32.3	43	0.7952	10%*			
	<i>ash-2</i> RNAi untr.	23	31.4	46					
	<i>ash-2</i> RNAi PQ	17	32.3	30	0.3873	26%*			
	<i>set-2</i> RNAi untr.	20	28.8	47					
	<i>set-2</i> RNAi PQ	23	28.7	39	0.3225	13%*			

Worms were treated with the indicated RNAi and sorted into L2^{ox} and L2^{red} at the F₂ generation (experiments 1–4).

[&]The corresponding empty-vector controls for experiments 1–4 are shown in Extended Data Table 1 experiments 7–10. Worms were treated with 1 mM paraquat for 10 h at the L2 stage following RNAi treatment with *ash-2* or *set-2* (experiments 5 and 6). P values for Kaplan–Meier survival analysis were calculated on the basis of the log-rank (Mantel–Cox) method. Samples were compared with the respective L2^{ox} or untreated control. Maximum lifespan was defined as the lifespan of the last 10% of the population.

*Percentage change based on mean lifespan.

[†]Percentage change based on maximum lifespan.

[‡]Grx1-roGFP2 ratio of L2^{ox} and L2^{red} subpopulations, based on BioSorter analysis during sorting. Data are mean ± s.e.m. n, number of worms. Two-sided unpaired t-test.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

BioRad - Image Lab (v5.2.1) was used for collecting gel and blot images
Union Biometrika - FlowPilot v. 1.5.9.4 was used to collect and analyze Biosorter data
No software or computer code was used to generate data

Data analysis

Prism - GraphPad Software Inc. (v7.0c)
Fiji - ImageJ (v1.52a) bundled with Java 1.8.0_112
Microsoft Office 2010 Excel
FastQC (version v0.11.3)
TopHat (version 2.0.13)
Bowtie2 (version 2.2.1.)
Cufflinks/CuffDiff (version 2.1.1)
CummeRbund R
Bioconductor Package GSA

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-sequencing data described in this study have been uploaded to the Gene Expression Omnibus (GEO) database with accession number GSE138502 and will be

made public upon publication.

Source data files as well as uncropped gel and blot images are provided with the paper. Supplementary Table 1 contains RNAseq data and analysis.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculations were performed. Sample size was determined to be adequate based on the magnitude and consistency of measurable differences between groups in preliminary experiments. RNA sequencing was performed in 4 biological replicates. The very high correlation between these replicates suggests that this is sufficient. In many experiments numerous worms were available and large sample sizes were used.
Data exclusions	In lifespan assays, we excluded plates based on the following pre-established criteria: 1) growth of bacteria that were not originally seeded on the plate 2) fungal growth 3) plate desiccation.
Replication	All findings reported were reliably reproduced in the lab on independent occasions : 1) different facilities 2) different C. elegans stocks 3) un
Randomization	Samples were allocated into groups (oxidized vs reduced) based on Grx1-roGFP2 readouts using a large - particle flow cytometer i.e. BioSorter
Blinding	Blinding was performed on lifespan assays

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	anti-H3 (Abcam, # ab1791; Lot #GR325252171-1; 1:2,000 working dilution) anti-H3K4me3 (Abcam, # ab8580; Lot # GR3275503-1; 1:1,000 working dilution) anti-H3K27ac (Abcam, # ab4729; Lot # GR3231988-1; 1:1,000 working dilution) anti-H3K27me3 (Millipore, # 07-449; Lot # 2919706; 1:1,000 working dilution) anti-ASH-2 (Abmart, # X3-G5EFZ3; Lot 31284-1; 1:1,000 working dilution) anti-β-tubulin (Santa Cruz, # sc-5274); Lot # B2218; 1:2,000-2,500 working dilution anti-ASH2L (Bethyl laboratories, # A300-489A; Lot # 2; 1:1,000 working dilution) anti-MLL1 (Bethyl laboratories, # A300-374A; Lot # 5; 1:500 working dilution)
Validation	All the antibodies used in this work are commercially available and have been published/cited. https://www.citeab.com/antibodies/763778-ab1791-anti-histone-h3-antibody-nuclear-loading-con https://www.citeab.com/antibodies/763751-ab8580-anti-histone-h3-tri-methyl-k4-antibody-chi?des=E833EC8BFE0454AA https://www.citeab.com/antibodies/778149-ab4729-anti-histone-h3-acetyl-k27-antibody-chip-g?des=EE13B14650FDA8F3 https://www.citeab.com/antibodies/221356-07-449-anti-trimethyl-histone-h3-lys27-antibody?des=348CD7BFA0CC732E https://www.citeab.com/antibodies/835112-sc-5274-tubulin-antibody-d-10?des=517C27105F70CC7F https://www.citeab.com/antibodies/655121-a300-489a-ash2-antibody?des=32EFD81E08FF28C8 https://www.citeab.com/antibodies/654913-a300-374a-ml1-antibody?des=620F371A0BC1D965

The ASH-2 C. elegans antibody from Abmart (X3 -G5EFZ3) was tested against C. elegans treated with ash-2 RNAi and with purified recombinant C. elegans ASH-2.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HeLa EM-II cell line was a gift from JK Nandakumar (University of Michigan)
Authentication	STR
Mycoplasma contamination	HeLa EM-II cell line has been tested in the lab for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	All C. elegans strains used in this study are detailed in Methods
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not involve field-collected samples
Ethics oversight	No ethical approval or guidance is required.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Key role for CTCF in establishing chromatin structure in human embryos

<https://doi.org/10.1038/s41586-019-1812-0>

Received: 14 January 2019

Accepted: 16 October 2019

Published online: 4 December 2019

Xuepeng Chen^{1,7}, Yuwen Ke^{1,7}, Kelian Wu^{2,7}, Han Zhao^{2,7}, Yaoyu Sun^{1,3}, Lei Gao¹, Zhenbo Liu¹, Jingye Zhang², Wenrong Tao², Zhenzhen Hou², Hui Liu², Jiang Liu^{1,3,4*} & Zi-Jiang Chen^{2,5,6*}

In the interphase of the cell cycle, chromatin is arranged in a hierarchical structure within the nucleus^{1,2}, which has an important role in regulating gene expression^{3–6}. However, the dynamics of 3D chromatin structure during human embryogenesis remains unknown. Here we report that, unlike mouse sperm, human sperm cells do not express the chromatin regulator CTCF and their chromatin does not contain topologically associating domains (TADs). Following human fertilization, TAD structure is gradually established during embryonic development. In addition, A/B compartmentalization is lost in human embryos at the 2-cell stage and is re-established during embryogenesis. Notably, blocking zygotic genome activation (ZGA) can inhibit TAD establishment in human embryos but not in mouse or *Drosophila*. Of note, CTCF is expressed at very low levels before ZGA, and is then highly expressed at the ZGA stage when TADs are observed. TAD organization is significantly reduced in CTCF knockdown embryos, suggesting that TAD establishment during ZGA in human embryos requires CTCF expression. Our results indicate that CTCF has a key role in the establishment of 3D chromatin structure during human embryogenesis.

During mouse embryogenesis, TADs are disorganized in 2-cell embryos and become increasingly established during development^{7,8}. To study the process of 3D chromatin structure reprogramming in human embryos, we examined human sperm, 2-cell embryos, 8-cell embryos, morula, blastocysts and six-week-old embryos. We optimized a previous low-input Hi-C (genome-wide chromosome conformation capture) method⁸ by using as few as 50–100 cells to construct Hi-C libraries in human early embryos (Methods). Our new ultra-low-input Hi-C method can detect very similar chromatin structures in a single mouse blastocyst to those obtained from pooled mouse blastocysts using the previous method⁸ (Extended Data Fig. 1a–c). At least two biological replicates were sequenced for each collected developmental stage, (Supplementary Table 1) and showed high reproducibility (Extended Data Fig. 1d–f).

TAD structures in human early embryos

Next, we examined chromatin interactions in human sperm and embryos. We did not detect the characteristic ‘triangle’ interactions of TAD structures in human 2-cell embryos. There was a low level of these interactions in 8-cell embryos, and the level increased during human embryonic development (Fig. 1a). To rule out the possible influence of read depth on the analysis, we picked the same number of reads at random when plotting interaction heat maps for each stage, showing consistent results (Extended Data Fig. 2a).

We further investigated TAD reprogramming during human embryogenesis. We used a TAD separation-score method (Methods) to call TAD domains and TAD boundaries (Supplementary Table 2). The majority of TAD domains can be detected even with very low read depth (Extended Data Fig. 2b, c). We also calculated TAD signal and directional index⁹. Our data show that TAD signal variance and directional index are lowest at the 2-cell stage and increase gradually through development (Fig. 1b, Extended Data Fig. 2d, e). To exclude possible experimental bias in TAD analysis of human embryos, we used mouse morula embryos as spiked-in controls, which were mixed in with human 2-cell, 8-cell and morula embryos. We then constructed Hi-C libraries for these mixed samples (Methods). In parallel, we also generated Hi-C libraries for mouse morula embryos that were not previously mixed with human samples. These results show that TAD structures from the mixed samples remain obscure in human 2-cell embryos and become clearer in 8-cell embryos and morula embryos (Extended Data Fig. 2f, g), whereas all the spiked-in mouse morula embryos show clear TAD structures (Extended Data Fig. 2h). TAD signal analysis from the mixed samples supports our finding that TAD structures become established during human embryogenesis (Extended Data Fig. 2i).

In sum, these data show that TAD structures are largely absent in human 2-cell embryos, are weakly present in 8-cell embryos, and become increasingly evident during embryonic development.

¹CAS Key Laboratory of Genome Sciences and Information, Collaborative Innovation Center of Genetics and Development, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China. ²Center for Reproductive Medicine, Shandong University, The Key laboratory of Reproductive Endocrinology, Ministry of Education, Shandong University, Jinan, China.

³University of Chinese Academy of Sciences, Beijing, China. ⁴CAS Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. ⁵Center for Reproductive Medicine, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. ⁶Shanghai Key Laboratory for Assisted Reproduction and Reproductive Genetics, Shanghai, China. ⁷These authors contributed equally: Xuepeng Chen, Yuwen Ke, Kelian Wu, Han Zhao. *e-mail: liuj@big.ac.cn; chen zijiang@hotmail.com

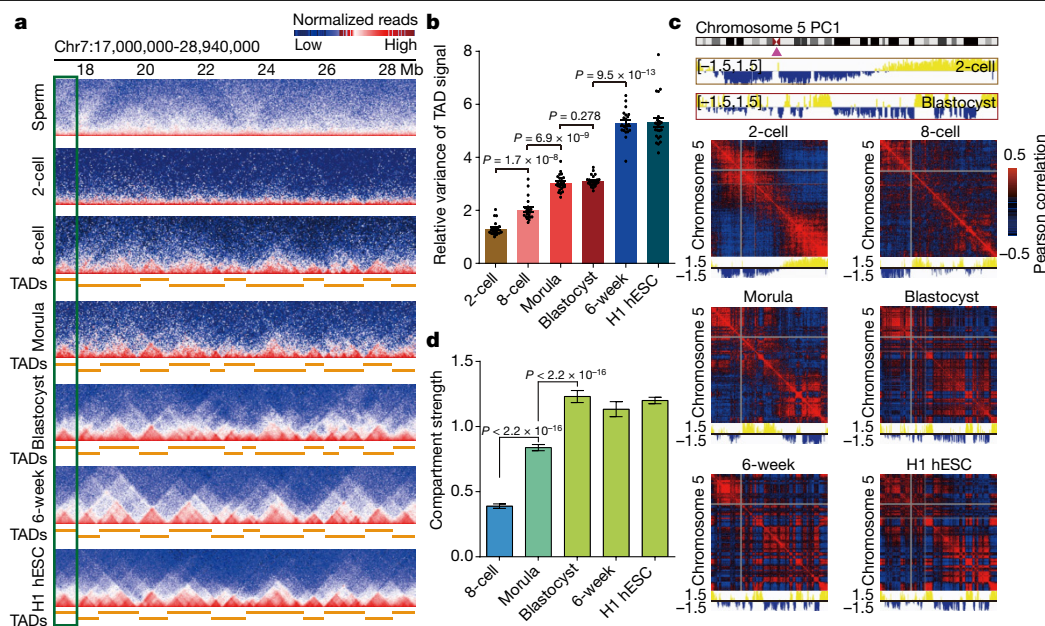


Fig. 1 | Three-dimensional chromatin structures of human sperm and embryos. **a**, Snapshot of the interaction heat map for high-order chromatin structures in human sperm, human embryos and H1 human ES cells (hESCs)¹¹ at 40-kb resolution (pooled biological replicates; $n = 2-3$). **b**, Relative variance of TAD signal in human embryos calculated with equal numbers of reads (generated from 2–3 biological replicates for each stage). Each dot represents one chromosome. P values are also shown; two-sided Wilcoxon rank-sum test.

The dynamics of A/B compartments

A/B compartmentalization is an important hierarchical structural feature of chromatin organization^{10,11}. We found that the chromosomes of human 2-cell embryos lost A/B compartmentalization, but usually partitioned into a few large interaction domains (Fig. 1c, Extended Data Fig. 3a, b), in contrast to those of mouse zygotes and early embryos (Extended Data Fig. 3c). This pattern is similar to the mouse inactive X chromosome¹². A/B compartmentalization is very weak at the human ZGA stage, and becomes more evident during development (Fig. 1c, Extended Data Fig. 3a). By contrast, there is clear A/B compartmentalization at the mouse ZGA stage (Extended Data Fig. 3c–e). The dynamic process of A/B compartmentalization in human embryos is further supported by compartment strength (Fig. 1d). Additionally, principal components (PC) analysis of the morula stage onwards showed bimodal distribution patterns of PC1 values, which are not observed at the 8-cell stage—suggesting poor compartment segregation at the 8-cell stage (Extended Data Fig. 3f). Together, these results suggest that A/B compartmentalization of chromosomes is not present at the 2-cell stage, is very weak at the 8-cell stage, emerges clearly at the morula stage and becomes more evident at the blastocyst stage.

Next, we investigated A/B compartment switches during human embryogenesis. We found that 13% of genomic regions show A/B compartment switches from the morula stage to the 6-week stage and 77% of them change their compartment statuses only once (Extended Data Fig. 4a). Then, we examined the potential effect of A/B compartment switches on embryonic development. Gene ontology analysis shows that genes that undergo A/B compartment switches are significantly enriched in developmental processes (Extended Data Fig. 4b, Supplementary Table 3). For example, genes undergoing a B-to-A switch between blastocyst and 6-week stage are enriched in telencephalon development and organ morphogenesis (Extended Data Fig. 4b). *NEFM* and *NEFL*, which function in maintaining the neuronal calibre, switch their B compartment status to A compartment and become highly expressed (Extended Data Fig. 4c).

Data are mean \pm s.e.m. **c**, PC1 value tracks and Pearson correlation heat maps for chromosome 5 at 500-kb resolution in human embryos with equal numbers of reads (generated from 2–3 biological replicates for each stage). **d**, Compartment strength in human embryos with equal numbers of reads. Data are mean \pm s.d., obtained by bootstrapping ($n = 100$). P values were calculated by one-sided t -test.

Intensive genome-wide epigenetic reprogramming occurs during human embryonic development^{13–15}. We analysed how A/B compartments and other epigenetic marks are orchestrated in human embryos. Consistent with the previous report in mouse embryos⁸, unmethylated CpGs in human embryos are enriched in A compartments; the level of CpG demethylation in A compartments is significantly higher than that in B compartments (Extended Data Fig. 4d, e). Furthermore, our data show that DNA demethylation levels are higher in regions that maintain their A compartment status from sperm to blastocyst (Extended Data Fig. 4f). This suggests that A compartments provide a more-open local nuclear environment that may enable easier access for DNA-binding proteins.

Human sperm do not have typical TADs

Previous studies have shown that TAD structures are present in mature mouse sperm^{8,16}. Surprisingly, we did not observe typical triangle TAD structures in human sperm (Fig. 1a). For example, there was no TAD boundary at the *HOXA* cluster region in human sperm, but it was present in human blastocysts (Fig. 2a) and mouse sperm (Extended Data Fig. 5a). To validate this observation, we plotted TAD signal variance at different read depths in human sperm and blastocysts. Unlike in mouse sperm⁸ and human blastocysts, the y-intercept of the human sperm line is close to 0 (Fig. 2b), suggesting an absence of TADs in human sperm. We further compared the density of interaction insert size between human sperm and mouse sperm. Our data show that human sperm presents one main peak around 4 Mb (middle-range), whereas mouse sperm has a shoulder peak around 933 kb and a long-range main peak at around 41 Mb (Extended Data Fig. 5b). Additionally, there was also a difference between human sperm and mouse sperm in the contact probability decay curve (Extended Data Fig. 5c). Together, these results suggest that human sperm do not contain TADs.

To exclude potential experimental bias, we mixed mouse sperm with human sperm and constructed a Hi-C library for the sperm mixture

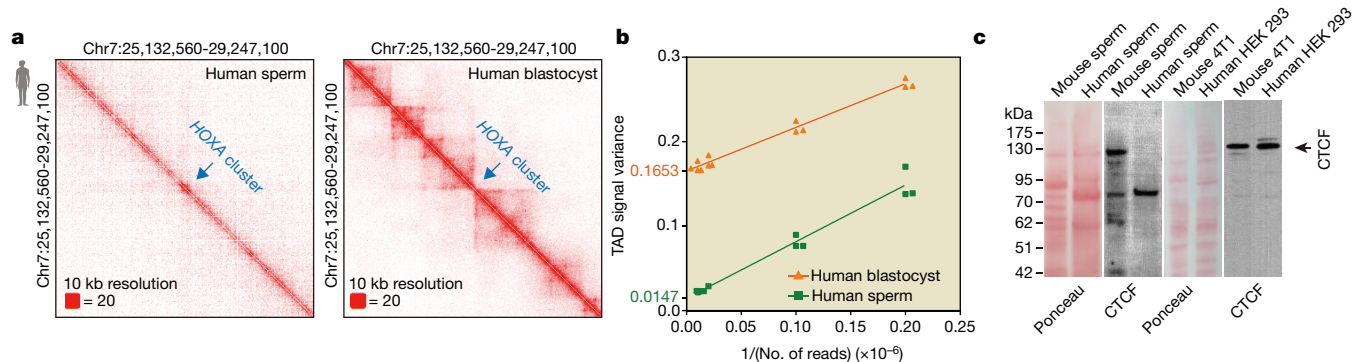


Fig. 2 | Human sperm does not have typical TAD structures. **a**, Interaction heat maps around the *HOXA* cluster in human sperm and blastocysts at 10-kb resolution (pooled biological replicates; $n = 3$). **b**, Linear regression curves for TAD signal variance versus read depth ($1/\text{no. of reads}$) in human sperm and blastocysts. The linear extrapolation of regression lines for human blastocysts

and sperm are marked on the y axis. Downsampling analysis at each read depth was performed three times ($n = 3$). **c**, Western blot for CTCF in mouse sperm, human sperm and somatic cell lines, with Ponceau staining. Black arrow, CTCF band. Experiment repeated on two biologically independent replicates per sample. For gel source data, see Supplementary Fig. 1.

(Methods). In parallel, we mixed human HeLa cells and mouse HT22 cells. Consistent with our previous result, we observed TADs from mouse sperm—but not from human sperm—in the sperm mixture (Extended Data Fig. 5d, e). By contrast, TAD structures were observed in both human HeLa cells and mouse HT22 cells (Extended Data Fig. 5d, e). These data corroborate the absence of TADs from human sperm.

We also inspected A/B compartmentalization in human sperm and found that chromatin from human sperm, both in isolation and when mixed with mouse sperm, shows clear A/B compartmentalization (Extended Data Fig. 5f).

The CTCF–cohesin complex has important roles in high-order chromatin structures^{4,17}. We investigated levels of CTCF and cohesin protein in human and mouse sperm. RAD21, a subunit of cohesin complex, is present in both human and mouse sperm (Extended Data Fig. 5g); however, CTCF was not detected in human sperm and was very weakly expressed in cell lines in which CTCF was depleted using short interfering RNA (siRNA) (Methods), but was detected in mouse sperm, human and control mouse cell lines (Fig. 2c, Extended Data Fig. 5h). Because depletion of CTCF can lead to the disruption of TAD structures¹⁸, the lack of CTCF may underlie the loss of TAD structures in human sperm.

Establishment of TAD boundaries

TAD structure is obscure in human 2-cell embryos. However, some regions in 2-cell embryos show evidence of insulator binding, which can separate upstream and downstream interactions (Extended Data Fig. 6a). At the later embryonic stages, most of these regions become TAD boundaries (Fig. 3a); these regions can therefore be regarded as immature TAD boundaries. In this study, we defined both immature TAD boundaries and mature TAD boundaries as insulated boundaries. We then analysed the dynamics of insulated boundaries during human embryonic development. Our data show that 635, 905, 317 and 306 insulated boundaries developed primarily at the 2-cell, 8-cell, morula and blastocyst stages, respectively (Fig. 3a, Supplementary Table 4). We also found immature TAD boundaries in mouse 2-cell embryos, and identified the stage-gained insulated boundaries in mouse (Extended Data Fig. 6b, Supplementary Table 5). We compared the insulated boundaries at the 2-cell stage with the total boundaries at the blastocyst stage in human samples, showing that 2-cell boundaries overlapped with 30% of blastocyst boundaries (Fig. 3b). Our data also show that the ZGA stage contained 67% of boundaries present in the blastocyst, which is similar to the proportion in mouse models (Fig. 3b). Furthermore, we found significant overlap when we compared human boundaries at the ZGA stage with those from mouse (Fig. 3c). For example, we found

insulated boundaries at the ZGA stage in both human and mouse around the *TTC1* and *CCNG1* genes (Extended Data Fig. 6c, d).

Next, we sought to identify the genomic regions that exhibit a preference for first forming insulated boundaries at an early stage. The distance of genomic regions that first gained boundaries at the 2-cell stage to housekeeping genes was smaller than that of boundaries regions that gained boundaries at a later stage (Fig. 3d). A similar result was observed in mouse embryos (Extended Data Fig. 7a). These data indicate that insulated boundaries gained at the earlier stage tend to locate around housekeeping genes in both human and mouse. We also found that the expression levels of housekeeping genes near the boundaries tended to be higher than those of other housekeeping genes (Extended Data Fig. 7b).

Repeat elements are reported to associate with TAD boundaries in cell lines⁹. We thus analysed the enrichment of repeat elements around stage-specifically gained boundaries in embryos. Our data illustrates that Alu repeats, but not LINE or MIR repeats (Fig. 3e, Extended Data Fig. 7c–e) repeats, are enriched around the insulated boundaries at the earlier stage in human. Similar results are observed in mouse embryos (Extended Data Fig. 7f–h). For example, insulated boundaries gained at the 2-cell stage around the human or mouse *RAB5A* gene are established at an Alu-dense region (Extended Data Fig. 7i). Moreover, our data show that AluS elements are highly enriched around insulated boundaries gained at the 2-cell stage (Fig. 3e). Moreover, AluS repeats around insulated boundaries gained at the 2-cell stage were highly expressed at the cleavage stage compared with other stages (Extended Data Fig. 7j, k). Collectively, these results show a preference for insulated boundaries to locate around Alu-dense regions in human embryos.

TAD establishment depends on ZGA

Previous reports show that TAD establishment is independent of ZGA in mouse and *Drosophila* embryos^{7,8,19}. We investigated whether these characteristics are conserved in human. We treated human zygotes with α -amanitin to inhibit ZGA (Extended Data Fig. 8a, b) and collected embryos at the 8-cell stage (Methods, Supplementary Table 6). Surprisingly, TAD structure was obscure in α -amanitin-treated 8-cell embryos (Fig. 4a, Extended Data Fig. 8c, d). Relative variance of the TAD signal in α -amanitin-treated embryos was also significantly lower than that in untreated 8-cell embryos (Fig. 4b). Therefore, TAD establishment in human embryos requires ZGA.

CTCF regulates the chromatin landscape

Next, we aimed to identify the protein involved in TAD establishment during ZGA. The cohesin complex and CTCF have important roles

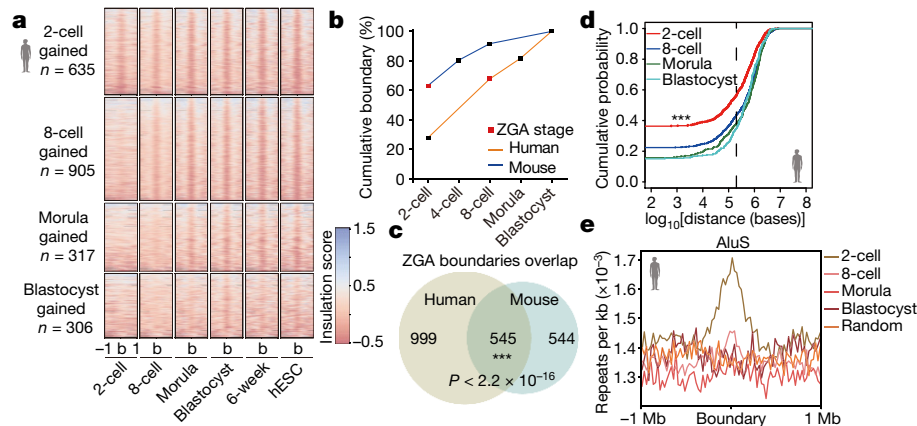


Fig. 3 | Establishment of insulated boundaries during human embryogenesis. **a**, Heat map for insulation score, centred at insulated boundaries gained at specific stages (± 1 Mb range). **n**, number of stage-specific gained insulated boundaries; **b**, insulated boundary centre. **b**, The cumulative percentage of insulated boundary number in each stage relative to the blastocyst boundary number. **c**, Venn diagram showing overlap between human ZGA boundaries and mouse ZGA boundaries (χ^2 test). **d**, Cumulative

distribution function plot for the distance of stage-specific gained boundaries to the closest housekeeping genes in human embryos (pooled data from 2–3 biological replicates). The dash line shows a distance of 200 kb. $P = 3.24 \times 10^{-8}$ (2-cell versus 8-cell stage); $P = 9.95 \times 10^{-8}$ (2-cell stage versus morula); $P = 3.83 \times 10^{-11}$ (2-cell stage versus blastocyst); two-sided Kolmogorov–Smirnov test. **e**, Enrichment of AluS elements at insulated boundaries gained at specific stages in human embryos.

in higher-order chromosome structures^{4,17}. Thus, we investigated differences in expression of these proteins^{20–22}. Subunits of cohesin complex, such as RAD21, are already highly expressed before ZGA in human embryos (Extended Data Fig. 9a). By contrast, CTCF expression is very limited before the ZGA stage, and is sharply increased at the 8-cell stage when TAD structure is first observed in human embryos (Fig. 4c, Extended Data Fig. 9b). In the α -amanitin-treated 8-cell embryos, CTCF expression is inhibited (Extended Data Fig. 9c). Consistently, immunostaining images show that CTCF protein is barely observed in the 2-cell nucleus (Fig. 4d). CTCF is clearly present in untreated 8-cell nucleus, but it is absent in the α -amanitin-treated

8-cell nucleus (Fig. 4d). These results suggest that CTCF expression requires human ZGA.

Next, we investigated the requirement of TAD establishment in human embryos for CTCF. We inhibited CTCF expression by injecting CTCF siRNA (siCTCF) into human zygotes, and collected embryos at the morula stage (Fig. 4d). Notably, triangle TAD structures were scarcely observed in the siCTCF morula (Fig. 4e, Extended Data Fig. 9d). Relative TAD signal variance supports inhibition of TAD establishment in the siCTCF morula (Extended Data Fig. 9e). Consistently, most of the TAD boundaries vanish in the control morula and become weaker in the siCTCF morula (Extended Data Fig. 9f). Thus, our data suggest

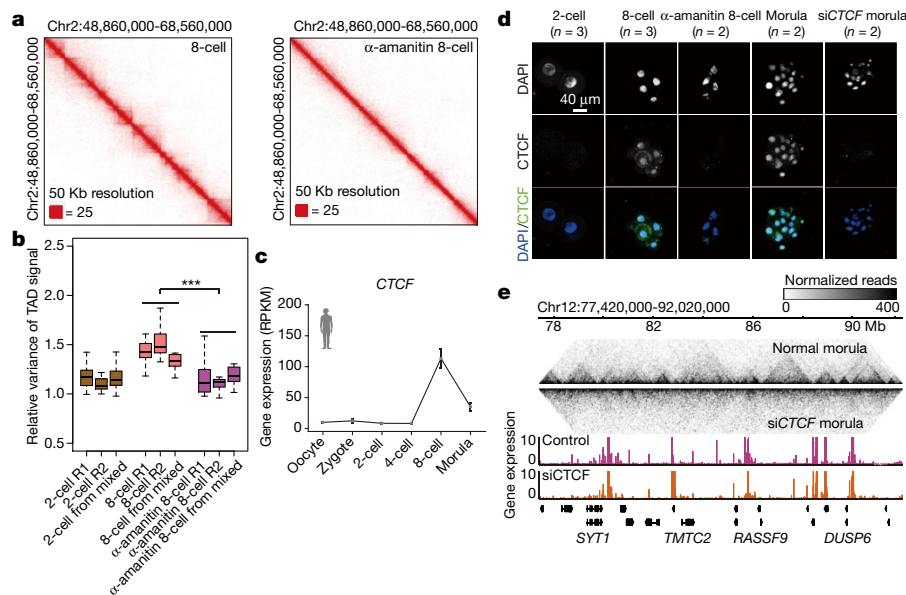


Fig. 4 | CTCF regulates the establishment of chromatin landscape in human embryos. **a**, Interaction heat maps in human 8-cell and α -amanitin-treated 8-cell embryos. **b**, Box plot for relative variance of TAD signal in human 2-cell ($n = 3$), 8-cell ($n = 3$) and α -amanitin-treated 8-cell embryos ($n = 3$). Boxes show 25th, 50th and 75th percentiles and whiskers show $1.5 \times$ the interquartile range. ***Adjusted $P < 0.001$ for all pairwise comparisons between 8-cell and α -amanitin-treated 8-cell (two-sided Wilcoxon rank-sum test with Benjamini–

Hochberg multiple testing correction). **c**, Dynamics of CTCF expression during human embryonic development (expression data from ref. 22; 3–20 cells for each stage). Data are mean \pm s.e.m. RPKM, reads per kilobase of transcript per million mapped reads. **d**, Immunofluorescence of CTCF in human embryos ($n = 2–3$). Scale bar, 40 μ m. **e**, Track snapshot for TAD structures in untreated control morula and siCTCF morula, overlaid with gene expression.

that CTCF expression during ZGA is required for TAD establishment in human embryos.

Frequently interacting region (FIRE) is a recently defined hierarchy of chromatin high-order structure²³. We identified FIREs at 40-kb resolution in human embryos. Human embryo FIREs are enriched in A compartments (Extended Data Fig. 10a–c). Our data show that both assay for transposase-accessible chromatin (ATAC) peaks²⁴ and DNase I hypersensitive sites (DHSs)²⁵ are enriched in FIREs (Extended Data Fig. 10d, e). To check whether CTCF knockdown has an effect on FIRE scores in human embryos, we analysed the FIRE score in the siCTCF morula and found that FIRE scores were significantly decreased (Extended Data Fig. 10f), consistent with previous reports in cell lines²⁵.

Additionally, we analysed the gene expression profile of the siCTCF morula and identified 565 downregulated genes in the siCTCF morula (Extended Data Fig. 10g, h). Our data show that highly downregulated genes in the siCTCF morula usually have CTCF-binding peaks around transcription start sites (TSSs) in human embryonic stem (ES) cells (Extended Data Fig. 10i).

Finally, we wanted to know whether CTCF expression is the sole determinant of TAD establishment. To address this question, we recovered CTCF expression by injecting CTCF mRNA into α -amanitin-treated human zygotes and collected the treated embryos at the 8-cell stage (Extended Data Fig. 11a, Methods). We did not observe clear triangle TAD patterns in 8-cell embryos that had been treated with α -amanitin and CTCF mRNA, in contrast to the patterns observed in untreated human 8-cell embryos and spiked-in mouse morula embryos (Extended Data Fig. 11b–e). Further insulation boundary analysis and TAD strength analysis²⁶ validated that TAD was not established in embryos treated with α -amanitin and CTCF mRNA (Extended Data Fig. 11f, g). In summary, our data suggest that CTCF expression is required, but is not the only factor needed, for the establishment of chromatin structure during human ZGA.

Although both human and mouse embryos show genome-wide reprogramming of high-order chromatin structure, there are substantial differences in chromatin structure between human and mouse embryos. Our data provide a valuable resource and mechanistic insights into the establishment of chromatin structure during mammalian embryonic development.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1812-0>.

- Fullwood, M. J. et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
- Atlati, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
- Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
- Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
- Hsieh, T. H. et al. mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108–119 (2015).
- Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
- Ke, Y. et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* **170**, 367–381 (2017).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
- Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
- Li, C. et al. DNA methylation reprogramming of functional elements during mammalian embryonic development. *Cell Discov.* **4**, 41 (2018).
- Smith, Z. D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
- Guo, H. et al. The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
- Jung, Y. H. et al. Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. *Cell Rep.* **18**, 1366–1382 (2017).
- Baranello, L., Kouzine, F. & Levens, D. CTCF and cohesin cooperate to organize the 3D structure of the mammalian genome. *Proc. Natl Acad. Sci. USA* **111**, 889–890 (2014).
- Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
- Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228 (2017).
- Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
- Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERV1/HERV1 retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
- Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
- Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
- Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
- Gao, L. et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell* **173**, 248–259 (2018).
- Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Human subjects

The regulatory framework about the use of human gametes and embryos for this research was based on the policies of the Human Biomedical Research Ethics Guidelines (set by National Health Commission of the People's Republic of China on 1 December 2016), the 2016 Guidelines for Stem Cell Research and Clinical Translation (issued by the International Society for Stem Cell Research, ISSCR) and the Human Embryonic Stem Cell Research Ethics Guidelines (set by China National Center for Biotechnology Development on 24 December 2003). These policies and guidelines permit human gametes, and/or human embryos created or genetic manipulated in vitro no more than 14 days, can be used specifically for scientific researches. All the experiments in this study are in compliance with these relevant ethical regulations.

The aim and protocols of this study has been reviewed and approved by the Institutional Review Board of Reproductive Medicine, Shandong University and Beijing Institute of Genomics. The protocols include the injection of siRNA into human embryos. Oocytes donated from patients taking in-vitro fertilization treatments, were fertilized using donated sperm by intracytoplasmic sperm injection (ICSI). Written informed consent was obtained from all oocyte and sperm donors. The donor women are 25–38 years old with tubal-factor infertility and their partners have healthy semen. The embryos with high quality were randomly assigned to experimental groups without examination about the gender of embryos owing to the limited cell number. All embryos are collected under standard clinical protocols. The number of embryos used at each developmental stage used for Hi-C in this study is shown in Supplementary Table 1. All control and siRNA-injected embryos were cultured no more than 7 days and only used for molecular research analyses. No statistical methods were used to predetermine sample size.

Human embryos collection

We collected human 2-cell, 8-cell, morula and morphological AA grade blastocysts at appropriate time after in vitro fertilization (IVF). The six-week embryos (6-week) were obtained from IVF couples by transvaginal multifetal pregnancy reduction surgery at the six-week stage of pregnancy.

For 2-cell embryos, they were from immature metaphase I (MI) oocytes that were not used clinically. MI oocytes matured in vitro and were fertilized by donated sperm. These fertilized embryos were then cultured in G1.5 medium (Vitrolife) in a humidified atmosphere at 37 °C with 6% CO₂ in air. The 2-cell embryos were collected and vitrified around 27 h after routine fertilization. The embryo vitrification was carried out as previously described²⁷. In brief, the embryos were incubated in vitrification solution 1 consisting of 8% ethylene glycol and 8% dimethyl sulfoxide in Cryobase (10 mM HEPES-buffered medium containing 20 mg ml⁻¹ human serum albumin and 0.01 mg ml⁻¹ gentamicin) at room temperature for 11 min. After initial shrinkage, embryos with original volume were transferred into vitrification solution 2 (16% ethylene glycol, 16% dimethyl sulfoxide and 0.68 M trehalose in Cryobase) for 1–1.5 min. Embryos were finally transferred onto Cryotop strip in an extremely small volume of solution (<0.1 ml) and plunged into liquid nitrogen. After addition of the protective cover, the Cryotop was transferred into liquid nitrogen for storage.

The embryos of 8-cell, morula and blastocyst stages were from donated frozen embryos with signed informed consent. These embryos were vitrified at 3 day, 4 day and 5 day after routine in vitro fertilization, respectively. The vitrified embryos on Cryotop strip were thawed rapidly by taking them from the liquid nitrogen after removal of the protective cover, then immersed in 2.5 ml of 37 °C warming solution 1

(1M trehalose in Cryobase) for 1 min on a heated stage. Embryos were then transferred to 0.5 ml of warming solution 2 (0.5 M trehalose in Cryobase) for 3 min, and then placed into 0.5 ml Cryobase for 5 min followed by fresh 0.5 ml Cryobase for 1 min. Embryos were finally transferred to G1.5 or G2 medium (Vitrolife) to evaluate embryo quality.

The thawed embryos with high quality were picked randomly for experimental groups. The zona pellucida was removed by mechanical dissection with a glass needle. Embryos were washed several times by gentle pipetting with a narrow-bore glass pipette to remove the attached cumulus or polar bodies. After final washing with 0.1% BSA/PBS three times, the embryos were transferred into microcentrifuge tubes (Sorenson BioScience, 39640T). Experiments were performed on at least two biological replicates for each embryonic stage.

Ultra-low-input Hi-C

The Hi-C library generation with ultra-low number of cells was optimized according to previous protocols^{8,26}. Samples were fixed with 20 µl of freshly made 1% formaldehyde solution and incubated at room temperature for 10 min. To quench the reaction, 2.5 M glycine solution was added to a final concentration of 0.2 M. Samples were incubated at room temperature for 5 min and then were lysed directly in 100 µl of ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630) with a protease inhibitor cocktail on ice for 15 min. Samples were then centrifuged at 3,000g for 5 min and the supernatants carefully discarded. Pelleted nuclei were washed once with 100 µl of 1× NEBuffer 2; discard the supernatant and remain 4.5 µl, and add 0.5 µl of 5% sodium dodecyl sulfate (SDS). Tubes were gently tapped to mix the pellet and were incubated at 62 °C for 5 min. After incubating, 14.5 µl of water and 2.5 µl of 10% Triton X-100 were added to quench the SDS. Tubes were gently tapped to mix well, avoiding excessive foaming and then incubated at 37 °C for 15 min. Two and a half microlitres of 10× NEBuffer 2 and 10 U of MboI restriction enzyme (NEB, R0147) were added and chromatin was digested at 37 °C for 5 h. Samples were incubated at 62 °C for 20 min to inactivate MboI and then cooled to room temperature. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 5 µl of fill-in master mix (3.75 µl of 0.4 mM biotin-14-dATP, 0.45 µl of 10 mM dCTP/dGTP/dTTP mix, 0.8 µl of 5 U µl⁻¹ DNA polymerase I, large) was added. Samples were mixed by pipetting and incubated at 37 °C for 30 min. Ligation master mix (66.3 µl of water, 12 µl of 10× NEB T4 DNA ligase buffer, 10 µl of 10% Triton X-100, 1.2 µl of 10 mg ml⁻¹ bovine serum albumin, 1 µl of 400 U/µl T4 DNA ligase) was added and samples were incubated at 16 °C for 18 h. Nuclei were pelleted by centrifugation for 5 min at 3,000g and were washed with 100 µl of 10 mM Tris buffer, pH 8.0. Pellets were then resuspended in 50 µl of 10 mM Tris buffer with 2 µl of 20 mg ml⁻¹ proteinase K and incubated at 65 °C for 18 h. Proteinase K was inactivated by incubation at 75 °C for 30 min. To fragment the DNA ligation products, 50 ng of λ DNA was added, and the DNA samples were digested with 1 U AluI at 37 °C for 4 h and inactivated at 65 °C for 30 min. Then the fragmented DNA was treated with the End Repair/dA-Tailing Module (NEB, E7442L) and Ligation Module (NEB, E7445L) following the operation manual. Samples were prepared for biotin pull-down by washing with 50 µl of 10 mg ml⁻¹ Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) in 100 µl of 1× Tween washing buffer (1× TWB: 5 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20). Samples were separated on a magnet and the solution was discarded. Ligation products were mixed with 83.5 µl of 2× binding buffer (2× BB: 10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 2 M NaCl) and centrifuged at 10,000g, 1 min. Beads were resuspended with supernatant and incubated at room temperature for 30 min with rotation to bind the biotinylated DNA to the streptavidin beads. Samples were again separated on a magnet and the solution was discarded. Beads were washed by adding 200 µl of 1× TWB and tubes were heated on a Thermomixer at 55 °C for 2 min with mixing. Beads were pelleted again using a magnet and the supernatant was discarded. This wash step was repeated twice. Beads were then resuspended in

Article

100 µl of 10 mM Tris buffer and transferred to a new tube. Beads were pelleted and the buffer discarded. Beads were resuspended in 20 µl of 10 mM Tris buffer. The Hi-C library was amplified for 12 cycles of PCR with Q5 master mix (NEB, M0492L) following the operation manual. PCR products were confirmed by analysing 1 µl of product using the FlashGel System (Lonza, 57063). PCR was continued with additional cycles until bright DNA bands were seen. A bottle of Agencourt AMPure XP beads (Beckman Coulter, A63881) was warmed to room temperature and gently shaken to resuspend the magnetic beads. One hundred microlitres of beads was added to 200 µl of diluted PCR product (0.5 volumes). Samples were mixed by pipetting and incubated at room temperature for 10 min. Beads were pelleted on a magnet and the clear solution was transferred to a new tube. Another 30 µl of beads was added to the clear solution (0.65× volume), mixed by pipetting and incubated at room temperature for 10 min. Keeping the beads on the magnet, samples were washed twice with 200 µl of 70% ethanol (freshly made) without mixing. Ethanol was then completely removed. Beads were left on the magnet for 5 min to allow the remaining ethanol to evaporate. DNA was eluted by adding 20 µl of ddH₂O, mixing by pipetting and incubating at room temperature for 5 min. After separating on a magnet, the solution was transferred to a new tube. DNA was then quantified and sequenced using an Illumina sequencing platform.

For human 2-cell, 8-cell, morula, blastocyst and 6-week embryos, we performed ultra-low-input Hi-C experiments. At least two replicates are prepared for each stage. For spike-in experiments, we performed ultra-low-input Hi-C experiments in human 2-cell mixed sample, human 8-cell mixed sample, human morula mixed sample, human α -amanitin-treated 8-cell mixed sample and human 8-cell (α -amanitin plus *CTCF* mRNA) mixed samples. In these spike-in experiments, the indicated stages of human embryos were mixed with six mouse morula embryos respectively.

Mouse sperm and morula embryos collection

Mouse sperm cells were collected from the cauda epididymis in 8-week old C57BL/6J male mice (Beijing Vital River Laboratory Animal Technology, stock number 213). The cauda epididymis was isolated and washed with 1× PBS three times, and cut into pieces and placed in a 5 ml 1× EmbryoMax Human Tubal Fluid (HTF) (EMD Millipore, MR-070-D) with 10% FBS (Gibco, 16140071) at a 6-well plate (Corning). Two cauda epididymis in one well. Sperms were incubated in HTF for 30 min in a CO₂ incubator at 37 °C to reduce somatic cells contaminants. Then, the supernatant with swimming active sperm was transferred to a new well without disturbing the settled somatic cells and low-activity sperm on bottom. The incubation–transfer cycle was repeated 4–5 times. The quality of samples was assessed under microscopy before snap freezing. The purity of sperm was over 99%. The sperms were then collected at minimal volume and frozen.

The morula embryos were from the cross of C57 BL/6J female and PWK/PhJ male. C57BL/6J female mice aged 4 to 6 weeks were intraperitoneally injected with pregnant mare serum gonadotropin (PMSG, 5UI) followed by human chorionic gonadotropin (hCG, 5UI) 48 h later. Morula embryos were flushed with M2 medium (Sigma) from the reproductive tract and the uteri at about 70–82 h after hCG administration. The morula embryos with high quality were confirmed with microscope. The embryos were serially washed to remove contaminants and were treated with acid Tyrode's solution (Sigma) for a few minutes to remove zona pellucida by pipetting through a fine pulled-glass needle. The embryos were then manually picked into PCR thin-wall tubes and stored at –80 °C.

All animal maintenance and experimental procedures were carried out according to guidelines of Institutional Animal Care and Use Committee (IACUC) of Beijing Institute of Genomics, CAS.

Sperm in situ Hi-C library generation

The generation of Hi-C libraries with sperm was optimized according to previous protocols³. Three million human sperm cells were used for each sample. For spike-in experiment, 3 million human sperm and

1–3 million mouse sperm were mixed together and use for one sample. As a control experiment, 3 million human cells HeLa (ATCC CCL-2; tested for mycoplasma contamination but not authenticated) and 3 million mouse cells HT22 (EMD Millipore, SCC129; tested for mycoplasma contamination but not authenticated) were mixed and used for one sample, and the Hi-C libraries generation of mixed somatic cell sample were followed the methods as previously published³.

Sperm cells were fixed with 1 ml of freshly made 1% formaldehyde solution and incubated at room temperature for 10 min. To quench the reaction, 2.5 M glycine solution was added to a final concentration of 0.2 M. Samples were incubated at room temperature for 5 min and then centrifuge for 5 min at 3,000g at 4 °C. Supernatant was discarded. The pellet was washed with ice-cold 1× PBS and spun for 5 min at 3,000g at 4 °C. Supernatant was discarded and cell pellets were flash-frozen in liquid nitrogen and stored at –80 °C until used, or used immediately. The sperm pellet was resuspended with 1 ml of ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630, 0.05% L- α -lysophosphatidylcholine) with a protease inhibitor cocktail. The cells were homogenized 5 times with a Dounce on ice and incubate for 30 min. Samples were then centrifuged at 3,000g for 5 min and the supernatants carefully discarded. Pelleted nuclei were washed once with 1 ml of ice-cold Hi-C lysis buffer. Supernatant was discarded and pellet was resuspended with 50 µl of 0.5% SDS and incubated at 62 °C for 10 min. After incubating, 145 µl of water and 25 µl of 10% Triton X-100 were added to quench the SDS. Tubes were gently tapped to mix well, avoiding excessive foaming and then incubated at 37 °C for 15 min. Twenty-five microlitres of 10× NEBuffer 2 and 100 U of MboI restriction enzyme (NEB, R0147) were added and chromatin was digested at 37 °C for overnight with rotation. Samples were incubated at 62 °C for 20 min to inactivate MboI and then cooled to room temperature. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 50 µl of fill-in master mix (37.5 µl of 0.4 mM biotin–14-dATP, 4.5 µl of 10 mM dCTP/dGTP/dTTP mix, 8 µl of 5 U µl^{–1} DNA polymerase I) was added. Samples were mixed by pipetting and incubated at 37 °C for 1.5 h. Ligation master mix (663 µl of water, 120 µl of 10× NEB T4 DNA ligase buffer, 100 µl of 10% Triton X-100, 12 µl of 10 mg ml^{–1} bovine serum albumin, 10 µl of 400 U µl^{–1} T4 DNA ligase) was added and samples were incubated at 16 °C for over 10 h with rotation. Nuclei were pelleted by centrifugation for 5 min at 3000g and were washed with 1× PBS. Pellets were then resuspended in 400 µl 1× PBS, and 16.7 µl of 20 mg ml^{–1} proteinase K and 40 µl 10% SDS were added, incubated at 55 °C for 30 min. The sample was added to 43.3 µl 5 M NaCl, 12 µl 0.5 M EDTA, 24 µl 1 M DTT, 50 µl 10% SDS and incubated at 68 °C overnight. The sample was cooled to room temperature, and 1,250 µl pure ethanol and 50 µl 3 M sodium acetate, pH 5.2 were added, mixed by inverting and incubated at –80 °C for over 1 h. Sample was centrifuged at 13,000g, 4 °C for 20 min. Tubes were kept on ice after spinning and the supernatant was carefully removed by pipetting. The pellet was washed with 800 µl of 70% ethanol twice and centrifuged at 13,000g for 5 min, dissolved in 100 µl of 1× Tris buffer (10 mM Tris-HCl, pH 8) and incubated at 37 °C for 15 min to fully dissolve the DNA. DNA was quantified by Qubit dsDNA High Sensitivity Assay (Life Technologies, Q32854). Biotin-14-dCTP at non-ligated DNA ends was removed with the exonuclease activity of T4 DNA polymerase (For 5 µg DNA: 1 µl 10 mg ml^{–1} BSA, 10 µl 10× NEBuffer 2, 1 µl 10 mM dNTP, and 5 U T4 DNA polymerase (NEB) in a total volume of 100 µl). The reaction was incubated at 12 °C for 2 h and stopped by adding 2 µl 0.5 M EDTA pH 8.0. All reaction products were combined. To make the biotinylated DNA suitable for high-throughput sequencing using Illumina sequencers, it was sheared to a size of 300–500 bp. Sheared DNA was transferred to a fresh 1.5-ml tube, and an equal volume of 2× binding buffer (2X BB: 10 mM Tris-HCl (pH 7.5); 1 mM EDTA; 2 M NaCl) was added, then centrifuged at 13,000g for 5 min. Sample was prepared for biotin pull-down by washing 50 µl of 10 mg ml^{–1} Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) with 400 µl of 1× Tween washing buffer (1× TWB: 5 mM Tris-HCl (pH 7.5);

0.5mM EDTA; 1M NaCl; 0.05% Tween 20), separated on a magnet and solution was discarded. Beads were resuspended with sheared DNA supernatant, incubated at room temperature for 30 min with rotation to bind biotinylated DNA to the streptavidin beads, separated on a magnet and solution was discarded. Beads were washed by adding 500 μ l of 1 \times TWB. The tubes were heated on a Thermomixer at 55 °C for 2 min with mixing. The beads were reclaimed the beads using a magnet and the supernatant discarded. The wash was repeated twice. Beads were resuspended in 100 μ l 1 \times Tris buffer (10 mM Tris-HCl, pH 8) and transferred to a new PCR tube. Beads were reclaimed and buffer discarded. Beads were resuspended in 50 μ l 1 \times Tris buffer (10 mM Tris-HCl, pH 8). Then, the beads binding with fragmented DNA were treated with the End Repair/dA-Tailing Module (NEB, E7442L) and Ligation Module (NEB, E7445L) following the operation manual. After ligation, the beads were washed with 1 \times TWB twice and 1 \times Tris buffer (10 mM Tris-HCl, pH 8). Beads were resuspended in 20 μ l of 1 \times Tris buffer (10 mM Tris-HCl, pH 8). The sperm Hi-C library was amplified for 10 cycles of PCR with Q5 master mix (NEB, M0492L) following the operation manual. PCR products were confirmed by analysing 1 μ l of product using the FlashGel System (Lonza, 57063). PCR was continued with additional cycles until bright DNA bands were seen. A bottle of Agencourt AMPure XP beads (Beckman Coulter, A63881) was warmed to room temperature and gently shaken to resuspend the magnetic beads. One hundred microlitres of beads was added to 200 μ l of diluted PCR product (0.5 \times volumes). Samples were mixed by pipetting and incubated at room temperature for 10 min. Beads were pelleted on a magnet and the clear solution was transferred to a new tube. Another 30 μ l of beads was added to the clear solution (0.65 \times volume), mixed by pipetting, and incubated at room temperature for 10 min. Keeping the beads on the magnet, samples were washed twice with 200 μ l of 70% ethanol (freshly made) without mixing. Ethanol was then completely removed. Beads were left on the magnet for 5 min to allow the remaining ethanol to evaporate. DNA was eluted by adding 20 μ l of ddH₂O, mixing by pipetting and incubating at room temperature for 5 min. After separating on a magnet, the solution was transferred to a new tube. DNA was then quantified and sequenced using an Illumina sequencing platform.

CTCF western blot

To confirm the CTCF or RAD21 antibody, 0.5 million HEK 293 cells (obtained from ATCC: ATCC CRL-1573; tested for mycoplasma contamination but not authenticated) and 0.5 million 4T1 cells (obtained from ATCC: ATCC CRL-2539; tested for mycoplasma contamination but not authenticated) were resuspended separately in 50 μ l 4% SDS and boiled for 5 min. To exclude the possibility of the CTCF antibody false positive detection on the alternate paralogue protein, we transfected HEK 293 cells and HeLa cells with human *CTCF* siRNA no. 1 (GGAGCCUGCCGUGAGAAUUTT)²⁸ and human *CTCF* siRNA no. 2 (CAGAGAAAGUGGUUGGUA)²⁹ or control oligonucleotides for two-rounds (denoted as si*CTCF* cell lines and control cell lines). The HEK 293 cells and HeLa cells were also resuspended separately in 50 μ l 4% SDS, boiled for 5 min. Protein concentration was assessed by BCA Protein Assay Kit (Cat no: CW0014, CWBIO). SDS-PAGE loading buffer (5 \times) (cat no: CW0027, CWBIO) was added into sample before loading to SDS-PAGE gel. The same total protein amount, 20 μ g, was loaded.

For mouse and human sperm samples, 10 million sperm cells were resuspended separately in 50 μ l 4% SDS, and sonicated using Diagenode Bioruptor sonication device for 10 cycles by 30 s on and 30 s off, then boiled for 30 min. Protein concentration was assessed by BCA Protein Assay Kit (cat no: CW0014, CWBIO). SDS-PAGE loading buffer (5 \times) (cat no: CW0027, CWBIO) was added into sample before loading to SDS-PAGE gel. The same total protein amount, 30 μ g, was loaded.

Samples were run on 4–12% SDS-PAGE and transferred to 0.2- μ m nitrocellulose membrane (Bio-Rad). Transfer was carried out in 25mM Tris base, 192 mM glycine, 20% methanol solution. The membranes

were stained with Ponceau red and photographed using a camera, then were blocked with 5% (w/v) milk buffer PBS/0.1% Tween 20 for 1 h at room temperature. The primary rabbit monoclonal antibody used was: CTCF (cat no: ab188408, Abcam) or RAD21 (cat no: ab992, Abcam) at a dilution of 1:1,000, 4 °C for overnight. Membranes were washed three times with TBST. The secondary antibody used was: goat anti-rabbit IgG, HRP conjugated (cat no: CW0103, CWBIO) at a dilution of 1:1,000, 1 h at room temperature. Membranes were washed three times with TBST. Signal development, followed the kit manufacturer's recommendations (cat no: CW0049, CWBIO). Excess reagent was removed and membrane was covered in transparent plastic wrap. Image was developed in a dark room for chemiluminescence or with digital imaging equipment.

ZGA block with α -amanitin treatment in human embryos

Zygotes were collected after IVF or ICSI. Zygotes were transferred to G1 media in the presence or absence of α -amanitin (25 ng μ l, Sigma-Aldrich). After the control and α -amanitin-treated embryos reached the 8-cell stage, embryos of high quality were collected. Others with fragments or those that were arrested at other stages were discarded. The zona pellucida was gently removed by mechanical dissection with a glass needle. The cells were washed three times with 1 \times PBS to avoid potential contamination. They were then prepared for Hi-C, RNA-seq and immunostaining.

CTCF knockdown by siRNA injection and CTCF overexpression by mRNA injection in human embryos

To investigate the function of CTCF in human embryonic TAD establishment, two independent siRNAs targeting human *CTCF* were used for microinjection: human *CTCF* siRNA no. 1 (GGAGCCUGCCGUGAGAAUUTT)²⁸; human *CTCF* siRNA no. 2 (CAGAGAAAGUGGUUGGUA)²⁹ and negative control siRNA (UUCUCCGAACGUGUCACGUGdTdT). For this assay, the donated IVM oocytes were used. They were in vitro-fertilized and cultured in G1.5 medium as described above. siRNA solution (10 mM) was loaded into injection pipette and injected into the zygote before pronuclear fading using Eppendorf PiezoXpert and Eppendorf CellTram vario microinjector. The injected embryos were cultured in G1.5 medium (Vitrolife) in a humidified atmosphere at 37 °C with 6% CO₂ in air. The injected embryos with normal morphology were collected at morula stage after fertilization. For RNA-seq library preparation, only one morula embryo was used for each assay. Morula embryos were collected for RNA-seq, Hi-C library generation and immunostaining, respectively. We collected three replicates in total for si*CTCF* morula: one replicate with human *CTCF* siRNA no. 1 (si*CTCF* morula no. 1) and two replicates with human *CTCF* siRNA no. 2 (si*CTCF* morula no. 2 and si*CTCF* morula no. 3).

To investigate whether human embryonic TAD establishment during ZGA is solely due to CTCF expression, we overexpressed CTCF in α -amanitin-treated embryos. The cDNA of complete open reading frame of human *CTCF* gene was cloned into the pCDNA3.1 vector. These plasmids were first linearized with NotI and mRNA was synthesized with SP6 polymerase using the mMessage mMachine kit (Ambion). The 1.5 μ g μ l⁻¹ mRNA was injected into the zygotes. Zygotes were transferred to G1 medium in the presence of α -amanitin (25 ng μ l⁻¹, Sigma-Aldrich). After these embryos reached the 8-cell stage, embryos with high quality were collected. Others with fragments or arrested at other stages were discarded. The zona pellucida was gently removed by mechanical dissection with a glass needle. The cells were washed three times with 1 \times PBS to avoid potential contamination. They were then prepared for Hi-C library generation and immunostaining.

SMART-seq2 library preparation

The human embryos were lysed directly and used for cDNA synthesis using SMART-Seq v.4 Ultra Low Input RNA Kit for Sequencing (Takara, 634888). In brief, the sample volume was adjusted to 9.5 μ l with nuclease-free water. After adding 1 μ l 10 \times reaction buffer (0.95 μ l 10 \times lysis buffer, 0.05 μ l RNase inhibitor), samples were incubated at room temperature for 5 min, then placed on ice. Next, 2 μ l of 3' SMART-Seq CDS

Article

Primer II A (12 mM) was added. Following incubation at 72 °C for 3 min, samples were placed on ice for 2 min. cDNA synthesis reaction was set up by adding 4 µl 5× ultra low first-strand buffer, 1 µl SMART-Seq v.4 Oligonucleotide (48 mM), 0.5 µl RNase Inhibitor (40 U ml⁻¹) and 2 µl SMARTScribed reverse transcriptase. The reaction was performed in a thermal cycler with following program: 42 °C for 90 min, 70 °C for 10 min, 4 °C hold. The first-strand cDNA product was amplified by adding 25 µl 2× SeqAmp PCR Buffer, 1 µl PCR Primer II A (12 mM), 1 µl SeqAmp DNA Polymerase and 3 µl nuclease-free water. Sixteen rounds of PCR amplification were used with the following program: 95 °C for 1 min; 98 °C for 10 s, 65 °C for 30 s and 68 °C for 3 min, repeat these 3 steps 15 times; 72 °C for 10 min; 4 °C hold. The amplified cDNA was purified using 1 volume SPRIselect beads, then fragmented to 200–400 bp using a Covaris sonicator (Covaris). Sequencing libraries were prepared with NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) as described above. To obtain an adequate amount of DNA for sequencing, the cycle of PCR amplification was determined according to the amount of 1 µl amplified DNA, which was evaluated using FlashGel System (Lonza, 57063). The libraries were sequenced on Hiseq X10 with paired-end 150 bp (Illumina).

Immunostaining

After removal of zona pellucida, human embryos were washed three times in 1×PBS, and fixed in PFA solution (4% paraformaldehyde, 0.04% Triton, 0.3% Tween in 1×PBS) for 20 min at RT. After washing three times in 1×PBS for 5 min, the embryos were permeabilized in 1×PBS containing 0.5% Triton for 20 min. Embryos were incubated with blocking buffer (3% BSA, 0.01% Tween-20 in 1×PBS) for 2 h at room temperature, then incubated with CTCF antibodies (cat. no.: ab188408, Abcam, 1:500 dilution) for about 12 h at 4 °C. After washing 6 times in 0.1% Triton in 1×PBS (PBST), embryos were blocked in blocking buffer again for 30 min and then incubated for 1 h at room temperature with Alexa Fluor 488-labelled goat anti-rabbit IgG (Beyotime, P0176). Embryos were washed in PBST 3 times, then stained with DAPI (Beyotime, C1002) for 10 min. After washing in PBST 3 times, embryos were mounted in Prolong Gold antifade reagent (Thermo Fisher, P10144). Confocal images were obtained by Zeiss LSM 710 confocal microscope using a 63× oil objective. All the staining assays were repeated independently at least twice. Images were processed and quantified by ZEN software (2010) and ImageJ software (v.1.52a).

Hi-C data processing

The HiExplorer suite^{30,31} (v.2.1) was used for the processing of valid mapped Hi-C reads. Ultra-low-input Hi-C data were first mapped to human genome hg19 using bwa mem with parameters ‘-E50 -L0’. Then, we filtered out read pairs which were not uniquely mapped (mapping score <15). Dangling end reads, same fragment reads, self-circled reads, self-ligation reads and other invalid Hi-C reads were also discarded. Details about data quality are summarized in Supplementary Table 1. After removing duplication, reads were used to generate raw Hi-C matrix at 10 kb, 40 kb, 100 kb and 200 kb resolution using hicBuildMatrix. In the raw contact matrix, rows and columns with zero or small total counts were removed because these bins were mostly from repetitive regions. After filtering low-count bins, the matrices were corrected by the ICE method³² using hicCorrectMatrix. We also converted valid read pairs into .hic format files with the juicer tool³³ pre command and use these .hic files for heat map visualization.

In addition, to avoid read depth influence on data interpretation, we also randomly picked up equal reads for each stage in the analysis when mentioned.

Reproducibility score

Reproducibility score is calculated from GenomeDISCO software³⁴. It uses random walks on the contact map graph for smoothing before comparing the interaction maps, resulting in a reproducibility score

that can be used for quality control of biological replicates. We calculated GenomeDISCO reproducibility score at 200-kb resolution for biological replicates.

TAD and TAD boundary identification

TADs and TAD boundaries (immature and mature TAD boundaries) were identified based on public TAD separation score method^{30,31} with hicFindTADs command in HiExplorer suite (v.2.1)^{30,31}. The TAD separation scores in human embryos were calculated at 40 kb resolution using parameters ‘-minDepth 300000 -maxDepth 3000000 -step 300000 -minBoundaryDistance 400000 -thresholdComparisons 0.01 -correctForMultipleTesting fdr -delta 0.01’.

The principle for TAD separation score and boundary detection is based on a graph clustering called the conductance of a cut. The conductance of a cut measure calculates the interactions between two clusters and the interactions within the clusters. It allows the identification of good clustering characterized by few interactions between clusters. It is similar to the TAD feature. Thus, TADs can be regarded as graph clusters of consecutive nodes (bins), and domain boundaries can be regarded as genomic positions that best separate two clusters (TADs).

The hicFindTADs command first transformed the Hi-C contact matrix into a z-score matrix $A = a_{ij}$. In the z-score matrix, each contact frequency is transformed into a z-score from the input contact matrix based on the distribution of all contacts at the same genomic distance. For an arbitrary bin l , the contacts between an upstream and downstream region of length w are all included in the z-score $A[\alpha_l, \beta_l]$ with row index $\alpha_l \in \{l - w, \dots, l\}$ and column index $\beta_l \in \{l, \dots, l + w\}$. For each matrix bin, we compute the TAD separation score (w) as the mean value of $A[\alpha_l, \beta_l]$. To reduce noise and improve robustness, a multi-scale version of the TAD-separation score is computed for different values of w that are averaged per bin from 300,000 to 30,000,000 with step 300,000 in an exponential size-growth way ($\text{minDepth} + (\text{step} \times \text{int}(x)^{1.5})$ for x in $[0, 1, 2, \dots]$). Genomic bins with a low TAD-separation score (local minima) are indicative of boundaries and stronger boundaries will have lower scores. For the local minima bin i , $i - v$ and $i + v$ are the bins at distance v upstream and downstream of i , respectively. We use the Wilcoxon rank-sum test to compare the values of $A[\alpha_i, \beta_i]$ with the values of the other two submatrices $A[\alpha_{i-v}, \beta_{i-v}]$ and $A[\alpha_{i+v}, \beta_{i+v}]$, respectively. The higher of the two P values is used. We corrected the P values using the false discovery rate (FDR) method, and regions with $q < 0.01$ were reported as boundaries. TADs were identified as regions between two boundaries.

Directional index and insulation score

Directional index⁹ can quantify the degree of upstream or downstream bias of a given bin. We calculated it for each bin based on the ICE-normalized and depth-normalized matrix as previously described⁸.

Insulation score was calculated as previously described¹² with the public code on Github (matrix2insulation.pl; <https://github.com/dekkerlab/giorgetti-nature-2016>). A sliding 480 kb × 480-kb square along the matrix diagonal was used. The IQR mean signal within the square was then assigned to the each 40-kb diagonal bin. This procedure was then repeated for all 480-kb diagonal bins. The insulation score was normalized relative to all of the insulation scores across each chromosome by calculating the log₂ ratio of each bin's insulation score versus the mean of all insulation scores. Valleys or minima along the normalized insulation score vector represent the loci of reduced Hi-C interactions that occur across the bin. These valleys or minima are interpreted as areas of high local insulation.

TAD structure strength

TAD structure strength was quantified as the ratio of the interactions within-TAD over the between-TAD intensity as previously described^{26,35}. For 90 × 90 contact enrichment map, the interactions within-TAD was the sum of the central square of the enrichment map, rows 30 to 59 and

columns 30 to 59 ([30:60, 30:60]) and the interactions in upstream and downstream neighbouring blocks was the between-TAD intensity, half of the sums of the regions [0:30, 30:60] and [30:60, 60:90]. TAD structure strength was both calculated for replicates and the merged.

TAD signal variance

TAD signal variance can indicate the strength of the TAD structure and calculated as previously described³⁶. We used intra-chromosomal maps at 40-kb resolution for each embryo stage with equal reads. In brief, TAD signal was calculated as the \log_2 ratio of the number of corrected upstream-to-downstream interactions within a 2-Mb window in each chromosome. Regions which contained less than 10 counts or gaps within a 2-Mb distance were filtered out. The variance was calculated for individual chromosomes. Comparisons between different stages were performed only for the bins shared by all of the different stages. Relative TAD signal variance was calculated relative to the minimum TAD signal variance. We also performed a downsampling analysis for human sperm and human blastocysts. Variance due to sampling noise was confirmed to be proportional to $1/\text{reads}$. The Wilcoxon rank-sum test was applied to test statistical significance.

Hi-C data mapping for mixed sample

After sequencing for the mixed samples, the human reads and mouse reads were separated from the mixture by using human genome reference and mouse genome reference. We mapped the mixed sample data to human hg19 genome and mouse mm10 genome with bwa mem with parameters '-E50 -L0', respectively. Separated data for human and mouse were processed as mentioned in the section 'Hi-C data processing'. In addition, we also combined the hg19 and mm10 genome together as reference genome. We mapped the reads to the combined reference genome. Those reads, with one end mapped to human and the other mapped to mouse genome, were all discarded.

A/B compartmentalization, compartment strength and A/B compartment dynamics

PC1 values from principal component analysis on correlation heat maps have been used to indicate the A/B compartment status¹⁰. We used HOMER³⁷ software with parameters '-res 100,000, -superRes 400,000' to obtain the PC1, PC2 and PC3 value and parameters '-res 500,000, -superRes 500,000' for equal read depth samples. Because sometimes PC2 or PC3 values reflect A/B compartments, we manually checked of PC1, PC2 and PC3 track with gene density and the plaid pattern in the correlation heat maps along each chromosome and got final 'PC1' list. The direction of the Eigen (PC) values is arbitrary, and therefore positive values were set to 'A' and negative values were set to 'B' based on their association with gene density.

Compartment strength was defined as the natural logarithm of the $AA \times BB / AB^2$ according to the public description²⁶. Enrichment contact maps of A/B interactions in Hi-C data were calculated at 500-kb resolution. A GC content profile was separated into 5 bins: minimum to 20th percentile, 20th percentile to 40th percentile and so on. For each pixel of the 5×5 matrix, O/E values were then calculated for loci belonging to each pair of bins as contact enrichment. GC content is known to highly correlate with the compartment profile. Thus GC content was used for compartment strength instead of PC1 values. We used the top 20% of GC content for A, and the bottom 20% GC content for B. For the error bar in evaluating the compartment strength in Fig. 1d, we obtained 100 5×5 compartment enrichment matrices by bootstrapping. For each pixel of the 5×5 compartment enrichment map, we took all the observed-over-expected values that contributed to this pixel and took a random sample with replacement of the same size that the contributing values. We then proceeded with downstream for each of the 100 reshuffled maps.

We identified A/B compartment dynamics regions as statistically significant variability in PC1 values across developmental stages

using analysis of variance with adjusted P value < 0.05 . Genes with A/B compartment switches were analysed for Gene Ontology (GO) enrichment with DAVID 6.8 (with parameters count ≥ 3 ; ease ≤ 0.05).

Density of paired loci at different interaction insert sizes

We examined the relative abundance of paired loci at different insert sizes in human sperm and mouse sperm as previously described⁸. We assigned the unique paired reads to restriction enzyme fragments and plotted the distribution of the \log_{10} of insert size between fragments.

Contact probability

Contact probability was calculated as previously described³⁶ and implemented with command line tool 'pairsqc.py' (<https://github.com/4dn-dcic/pairsqc>). It first divides distances into logarithmically spaced bins at \log_{10} scale at interval of 0.1 (growing by about 1.25-fold). For each bin, contact probability is computed as $\text{number_of_reads} / \text{number_of_possible_reads} / \text{bin_size}$. Number_of_reads is the number of interactions at corresponding distances. Number_of_possible_reads is computed as the sum of $L_{\text{chr}} - s_{\text{mid}} - 1$ over all chromosomes, where L_{chr} is the length of a chromosome. s_{mid} is the mid point of the bin at \log_{10} scale ($\text{bin } 10^{5.0} - 10^{5.1}$ has mid point $10^{5.05}$). Bin_size is computed as $\text{max distance} - \text{min distance}$ (for example, for bin $10^{5.0} - 10^{5.1}$, the binsize is $10^{5.1} - 10^{5.0}$).

Region comparison between human genome and mouse genome

For the region comparison between human and mouse, regions were shifted using the UCSC liftOver tool online for hg19 and mm10 genome with default parameters. For ZGA boundary overlap calculation, we lifted the mouse ZGA boundaries to human hg19 genome.

Identification of stage-specific gained boundaries

To analyse insulated boundaries between embryo stages, we first merged boundaries within 200 kb of each other across all examined embryo stages as common boundary reference (200-kb collapsed boundaries). It increased the robustness of boundary comparison. For stage-specific gained boundaries, we selected those boundaries that did not exist in previous stages, first emerged in the indicated stage and persisted until in blastocysts. For example, human 8-cell gained boundaries refer to this kind of boundary: they do not function as insulated boundaries in the 2-cell stage, and at the 8-cell stage they can be observed with evident insulation

Enrichment analysis of repeats on stage-specific gained boundaries

The repeat frequency in the genome was calculated for every 1-kb bin with bedtools coverage command. For generating the enrichment plots for repeats, the mid-point of each stage-specific gained boundary region was identified, and repeat frequency was calculated in 25-kb bins for ± 1 Mb from the boundary mid-point centre.

FIREs calling

FIREs calling was based on the previous method²³. For each 40-kb bin, the raw count was calculated as the total number of chromatin interactions within 200 kb. A normalization pipeline, 'HiCNormCis'²³, was applied to normalize raw count. A Poisson regression model was fitted for each 40-kb bin taking the raw count as the outcome variable, and three local genomic features, such as effective fragment length, GC content and mappability score, as the covariates. The residuals from the Poisson regression model were used as the HiCNormCis output. Next, for each 40-kb bin, z -score was calculated based on the HiCNormCis output. FIREs are defined as bins with a one-sided P value less than 0.05, corresponding to $-\ln(P)$ greater than 3. $-\ln(P)$ was chosen as the FIRE score.

DNA methylation data analysis

The human early embryos and sperm DNA-methylation data were downloaded from CRA000114³⁸, which had been deposited in the Genome

Article

Sequence Archive (GSA). Low-quality reads were removed by Trimmomatic v.0.32. The filtered reads were aligned to hg19 genome by Bismark with default parameters. PCR duplicates were removed by Picard and overlapped parts of pair-end reads were trimmed from one end by bamUtil. The methylation level of each CpG site was calculated by a custom script.

DNase-seq data analysis and ATAC-seq data analysis

DNase-seq data was analysed as previously described²⁵. Low-quality reads were removed and the remaining reads were cropped to 100 bp by Trimmomatic v.0.32. Paired reads and unpaired reads were used for mapping to human genome hg19 with Bowtie v.1.2.0³⁹ with parameter '-m 1'. Low mapping quality (MAPQ < 10) and PCR-duplicated reads were removed by SAMtools and Picard^{40,41}. DHSs were called by hot-spot algorithm with FDR < 0.01 (ref. ⁴²). DHSs covered by less than 8 reads were filtered out. DNase-seq signal tracks for visualization were generated by bamCoverage in DeepTools2 suite⁴³. ATAC-seq data were downloaded and analysed from GEO: GSE101571²⁴.

RNA-seq data processing

Human early embryo RNA-seq raw data were downloaded from GSE36552²². We used Refseq genes from the UCSC genome browser and Ensemble genes as gene annotation. Housekeeping gene annotation are downloaded from a previous publication⁴⁴. Low quality RNA-seq reads and adaptor sequences were removed by Trimmomatic v.0.32. Then reads were mapped to the hg19 genome by STAR (v.2.5.2b)⁴⁵ with default parameters. Gene expression FPKM was calculated by HOMER software. The ZGA genes were defined as those obviously upregulated in human 8-cell embryos (FPKM > 1 in 8-cell embryos and fold change > 3 comparing with 2-cell embryo)²⁵. To analyse the expression of Alu repeats in human early embryos, the total RNA-seq data were downloaded from GSE85632²¹. The reference repeat annotation was downloaded from rmsk in UCSC Table Browser. The FPKM value of each repeat or transposon was calculated by HOMER analyseRepeats.pl with parameters '-fpkm' and plotted as tracks in Extended Data Fig. 7j.

To find differential expression genes between α -amanitin 8-cell and control 8-cell or between siCTCF treatment and siControl morula, HOMER software command getDiffExpression.pl with -DESeq2 was employed. FDR < 0.05, |fold change| > 2 were used as cut-off for significant differently expressed genes.

RNA-seq tracks for visualization were generated by bamCoverage program in DeepTools2 with parameter '-normalizeUsingRPKM'.

Statistics and reproducibility

R and Prism were used for statistics analysis. Wilcoxon rank-sum test was used for statistical significance (Figs. 1b, 4b, Extended Data Figs. 2i, 4b, left, 4d-f, 7b, 8a, 9e, 10d, e) with the wilcox.test function in R. χ^2 test was applied for the statistical significance of the overlap between human ZGA boundaries and mouse ZGA boundaries relative to random (Fig. 3c) and the enrichment of FIREs in A compartments (Extended Data Fig. 10c). Two-sided Kolmogorov-Smirnov test were used for comparison the distance distribution of stage-gained insulated boundaries to the closest housekeeping genes in Fig. 3d and Extended Data Fig. 7a. One-sided *t*-test is used to calculate statistical significance for A/B compartment strength (Fig. 1d) and differential FIRE score (siCTCF - control) in FIREs and non-FIREs (Extended Data Fig. 10f). Fisher's Exact test was used for GO enrichment analysis by DAVID tool (v.6.8) with *P* < 0.05 cut-off in Extended Data Fig. 4b. For box plots, boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; NS, *P* > 0.05.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data generated for this study have been deposited to the Genome Sequence Archive with the accession number CRA000852 and CRA000108. Hi-C data of unmixed human sperm sample have been deposited in CRA000108, and the other data have been deposited in CRA000852. Raw image files used in the figures that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

Codes used for the analysis reported in this study are available at <https://github.com/ChenXP0310/2019-humanembryo3D>.

27. Roy, T. K., Bradley, C. K., Bowman, M. C. & McArthur, S. J. Single-embryo transfer of vitrified-warmed blastocysts yields equivalent live-birth rates and improved neonatal outcomes compared with fresh transfers. *Fertil. Steril.* **101**, 1294-1301 (2014).
28. Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996-1001 (2014).
29. Guillou, E. et al. Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* **24**, 2812-2822 (2010).
30. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
31. Wolff, J. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46** (W1), W11-W16 (2018).
32. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999-1003 (2012).
33. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95-98 (2016).
34. Ursu, O. et al. GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics* **34**, 2701-2707 (2018).
35. Gassler, J. et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600-3618 (2017).
36. Naumova, N. et al. Organization of the mitotic chromosome. *Science* **342**, 948-953 (2013).
37. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589 (2010).
38. Li, G. et al. Genome wide abnormal DNA methylome of human blastocyst in assisted reproductive technology. *J. Genet. Genom.* **44**, 475-481 (2017).
39. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
40. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498 (2011).
41. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
42. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264-268 (2011).
43. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160-W165 (2016).
44. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569-574 (2013).
45. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
46. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
47. Consortium, E. P.; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

Acknowledgements This work was supported by the grants from The Ministry of Science and Technology of China (2018YFC1004000 and 2018YFC1003300), National Natural Science Foundation of China (81430029, 91731312, 81871171, 31425015, 31630040, 31871454, 81622021 and 81871168), CAS funding (QYDZD-SSW-SMC016) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040200).

Author contributions J.L. and Z.-J.C. conceived the study. X.C., Y.K. and K.W. facilitated its designs. K.W., H.Z., J.Z., W.T. and H.L. collected human embryos. K.W. and Z.H. performed siRNA-microinjection in human embryos. Y.K., X.C. and Y.S. performed Hi-C library construction. Z.L. performed RNA-seq library construction. X.C., Y.S., and L.G. performed the bioinformatics analyses. X.C., Y.K., K.W., Z.-J.C. and J.L. interpreted the data. X.C., Y.K., Z.-J.C., and J.L. wrote the paper with the assistance of the other authors.

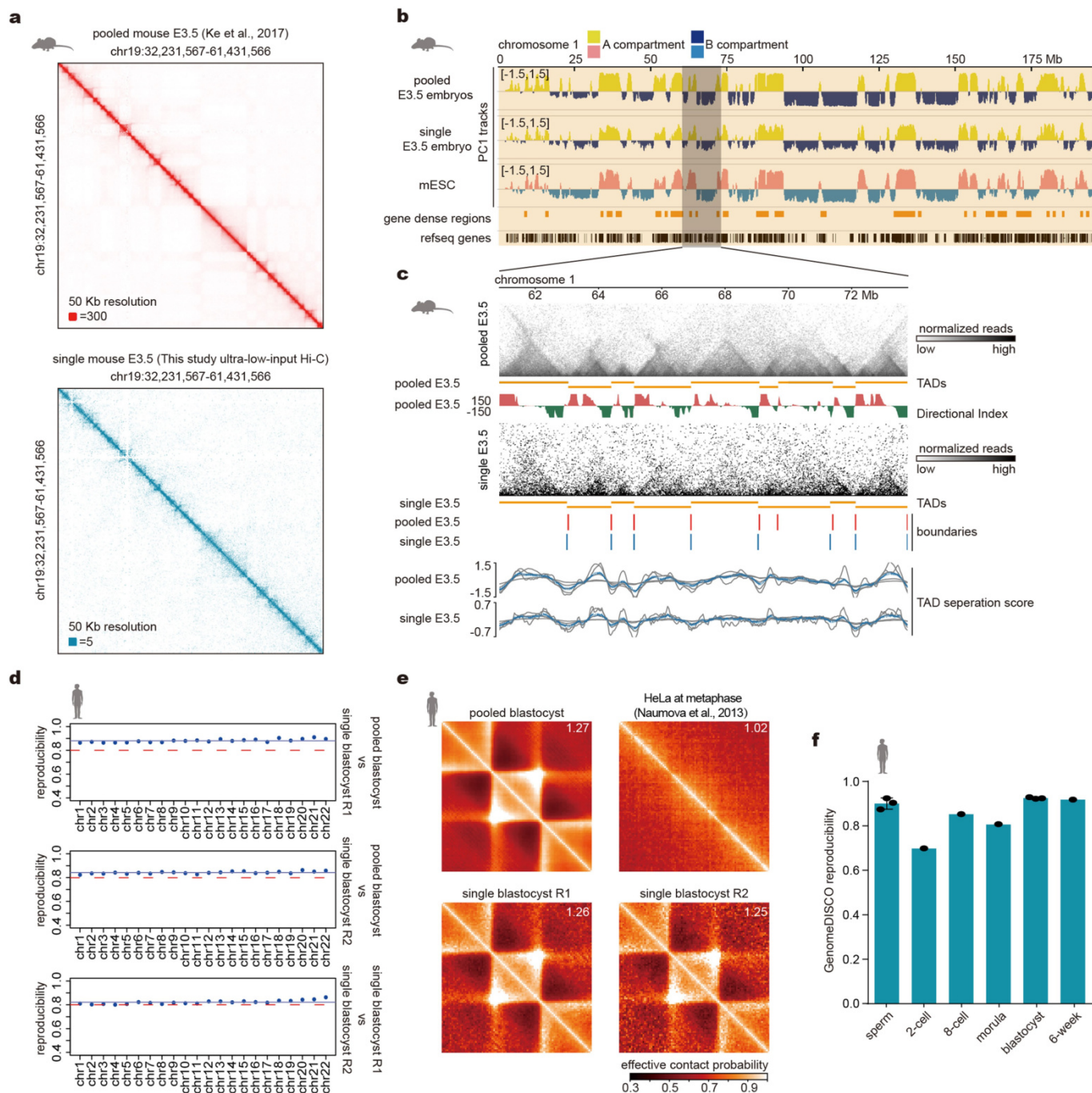
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1812-0>.

Correspondence and requests for materials should be addressed to J.L. or Z.-J.C.

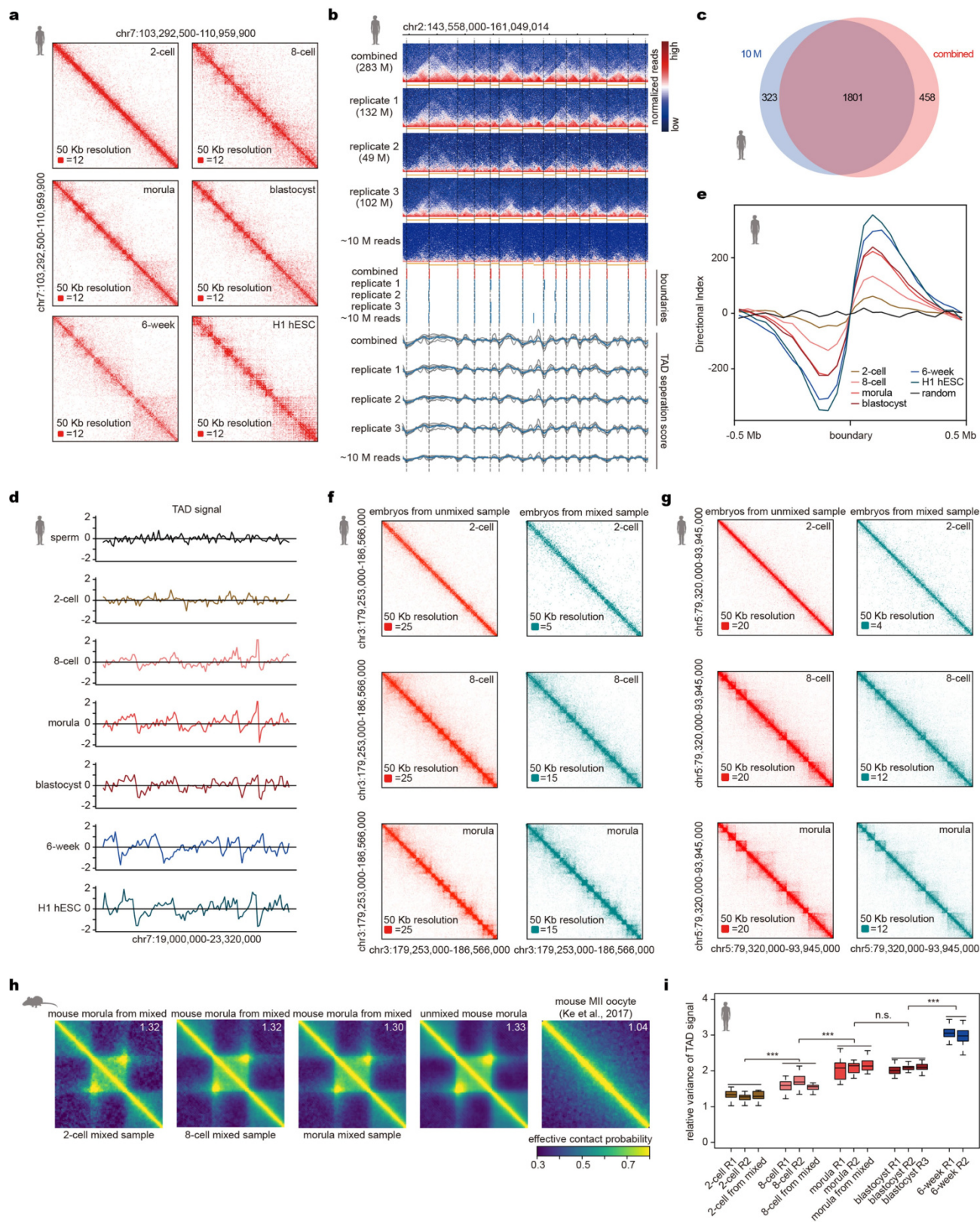
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Validation of optimized ultra-low-input Hi-C.

a, Interaction heat maps for mouse embryonic day (E)3.5 embryos. Top, pooled mouse E3.5 embryos Hi-C⁸. Bottom, a single mouse E3.5 embryo with ultra-low-input Hi-C from this study. **b**, A track snapshot of PC1 values for chromosome 1 in pooled mouse E3.5 embryos and single mouse E3.5 embryo along with public mouse ES cell (mESC) data⁹. **c**, Interaction heat maps for pooled mouse E3.5 embryos and single mouse E3.5 embryo for a zoomed-in region from chromosome 1 overlaid with TADs, directional index, boundaries and TAD separation scores. **d**, GenomeDISCO reproducibility (Methods) among the

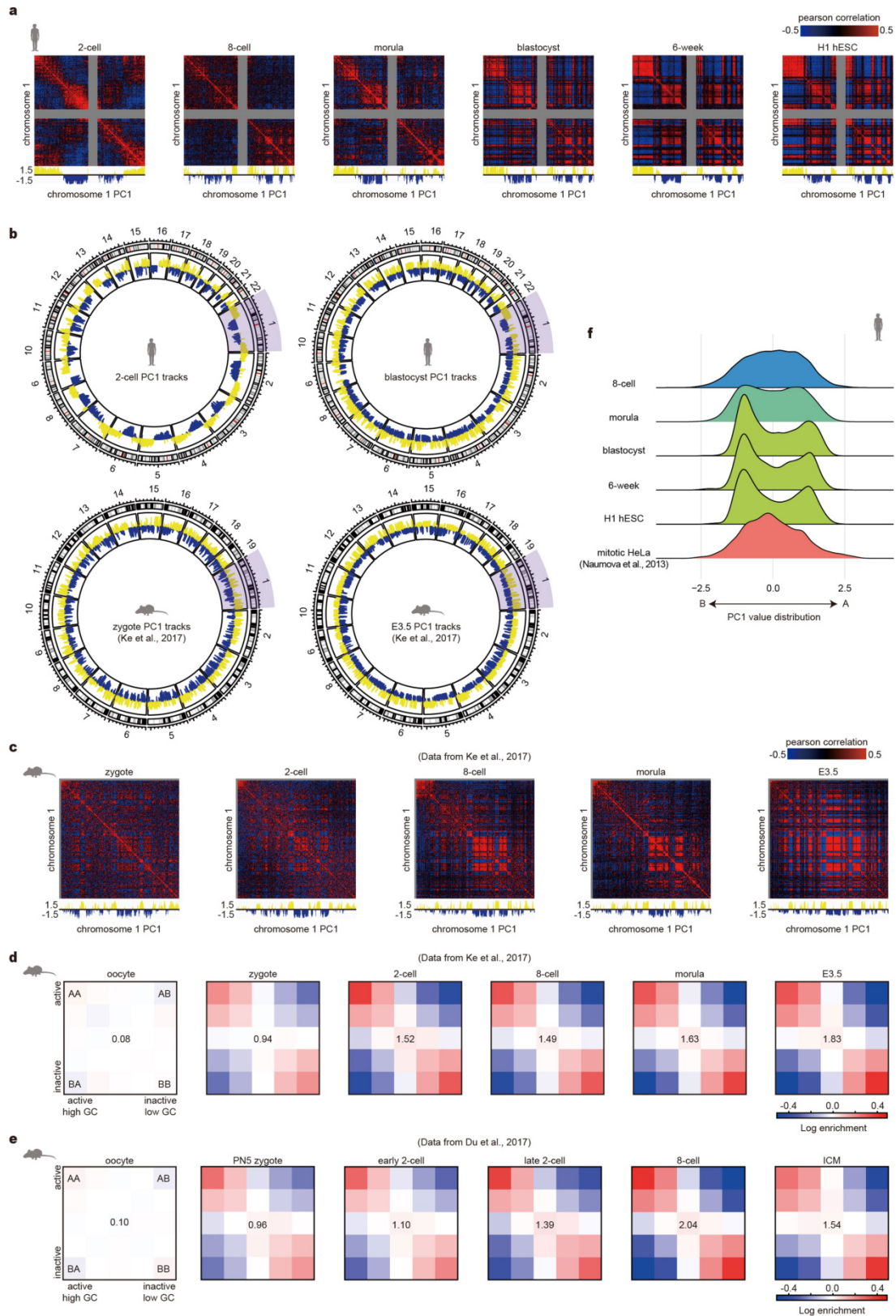
Hi-C data for each chromosome in pooled human blastocysts and two human single-blastocyst replicates. **e**, The strength of average TADs in pooled human blastocyst data, single human blastocyst data and metaphase HeLa data³⁶. TAD positions are annotated in combined human blastocyst data. TAD structure strength (Methods) is shown on each corresponding panel. **f**, GenomeDISCO reproducibility for biological replicates of human sperm and embryos. $n = 2-3$ independent biological replicates for each stage. Each dot represents one comparison between two biological replicates. Data are mean \pm s.d.



Extended Data Fig. 2 | See next page for caption.

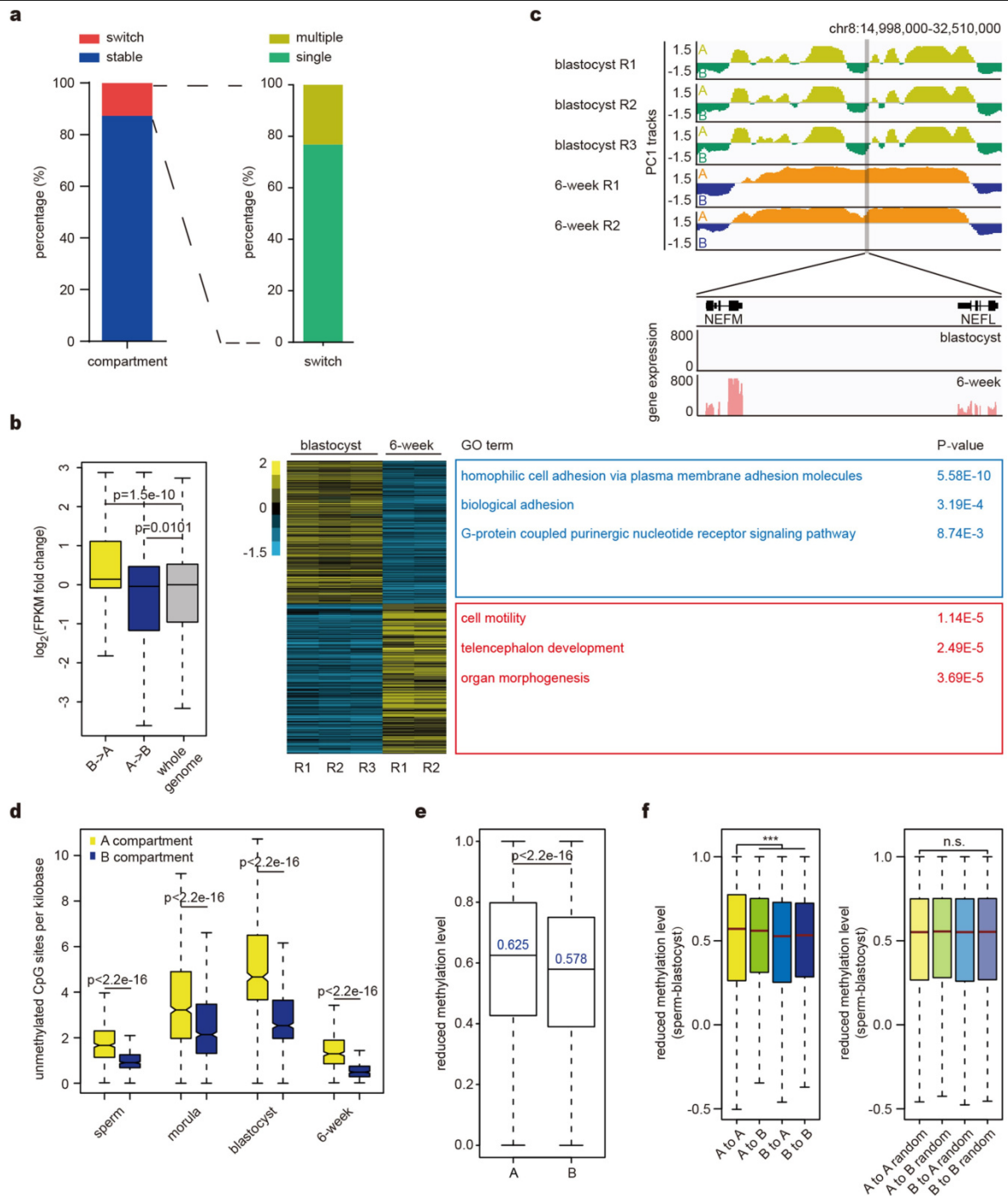
Extended Data Fig. 2 | TAD establishment during human embryonic development. **a**, Interaction heat maps for human embryos with equal reads along with human H1 ES cells. **b**, An example for TAD calling by using TAD separation score method in human blastocysts with different read depths. The orange track under each heat map represents TAD domains. The bottom panels are TAD boundaries and TAD separation score tracks. M, million reads. **c**, Venn diagram for the overlap of TAD boundaries of human blastocyst between 10 million read depth data (10 M, blue) and total combined data (total, red). **d**, An equal-read-depth TAD signal track snapshot for human sperm and embryos along with human H1 ES cells. **e**, Average directional index around human TAD boundaries (0.5 Mb) at different embryonic stages. Directional index generated by a random blastocyst valid read pair dataset is also shown as a control. **f, g**, Interaction heat map examples at 50-kb resolution for unmixed

human embryos and human embryos from mixed samples. **h**, The strength for average TADs in mouse morula embryos from mixed samples (six mouse morulae for each mixed sample), unmixed mouse morula and mouse MII oocyte. TAD positions are annotated in combined mouse E3.5 embryos. TAD structure strength is shown on each corresponding panel. **i**, Relative variance of TAD signal for replicates of human embryos including human embryos from mixed samples. We used the chromosome as the unit for this analysis. Two-cell ($n=3$), 8-cell ($n=3$), morula ($n=3$), blastocyst ($n=3$) and 6-week embryos ($n=2$). Boxes represent the 25th, 50th and 75th percentiles and whiskers show $1.5\times$ the interquartile range. ***adjusted $P < 0.001$ for all pairwise comparison between two stage samples (two-sided Wilcoxon rank-sum test with Benjamini-Hochberg multiple testing correction). NS, not significant.



Extended Data Fig. 3 | A/B compartmentalization dynamics during human embryonic development. **a**, Pearson correlation heat maps of chromosome 1 at 500-kb resolution in human 2-cell, 8-cell, morula and blastocysts, 6-week embryos and H1 ES cells (equal no. of reads generated from 2–3 biological independent replicates for each stage). **b**, Top, whole-genome PC1 value circle plots for human 2-cell embryos and human blastocysts (with R package *circize*⁴⁶). Bottom, PC1 tracks in mouse zygotes and mouse E3.5 embryos.

c, Pearson correlation heat maps at 500-kb resolution of chromosome 1 in mouse zygotes and early embryos⁸ (equal number of reads generated from 2–3 biological independent replicates for each stage). **d, e**, Five-by-five contact enrichment maps of A–B compartments averaged over genomic positions using GC content in mouse early embryo Hi-C data⁸ (d) and ref.⁷ (e). **f**, PC1 value distribution in human embryos and mitotic HeLa cells³⁶.

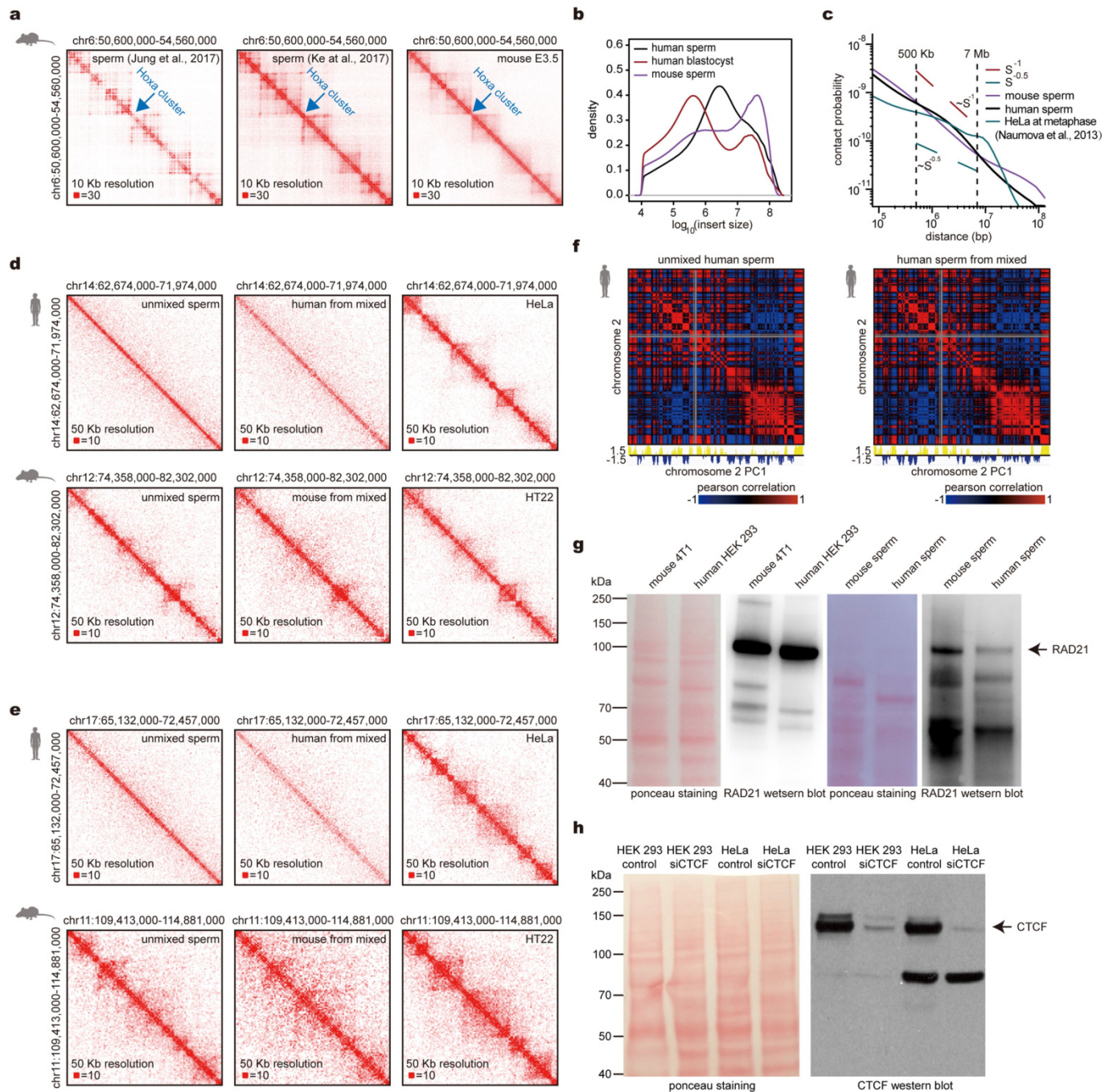


Extended Data Fig. 4 | A/B compartment switches during human embryonic development.

a, The percentage of genomic regions with A/B compartment switches from morula to 6-week stage, and the percentage of single A/B switched regions relative to total switched regions. **b**, Left, the box plot illustrates the gene expression dynamics with A/B compartment switches between human blastocysts and 6-week embryos (biological replicates pooled; blastocyst, $n = 3$; 6-week embryo, $n = 2$). P values are shown (one-sided Wilcoxon rank-sum test). Boxes represent the 25th, 50th and 75th percentiles and whiskers show $1.5 \times$ the interquartile range (**b**, **d-f**). Middle, PC1 heat map for A/B compartment switched regions. Heat map is sorted according to A/B compartment switched PC1 values. Right, GO enrichment for genes with A/B compartment switches between blastocysts and 6-week embryos using DAVID v.6.8. **c**, A/B compartment status around the *NEFM*-*NEFL* locus in human

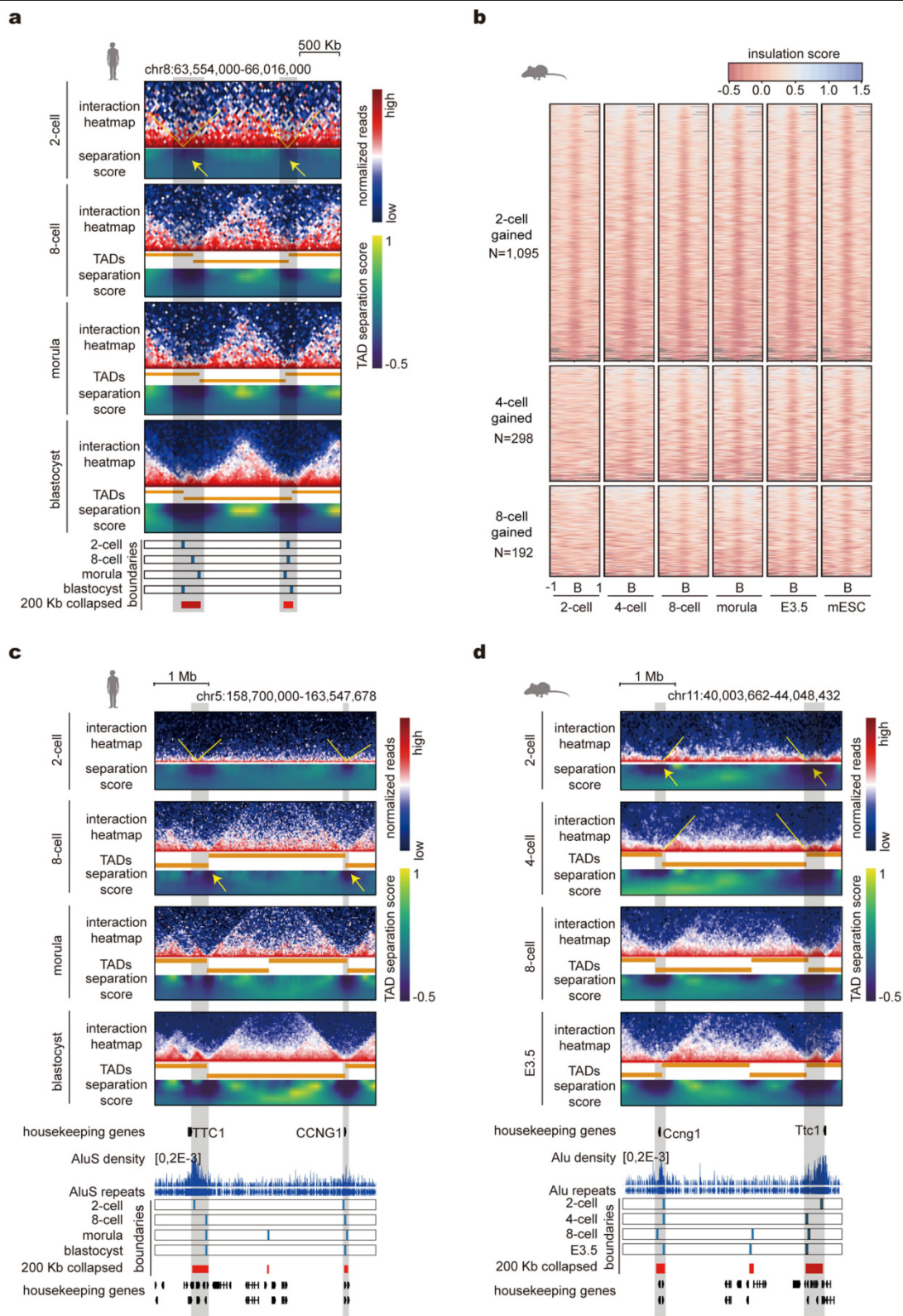
blastocysts and 6-week embryos overlaid with gene expression.

d, Unmethylated CpG (methylation level < 0.25) density within A or B compartments in human sperm and embryos (biological replicates pooled; sperm, $n = 3$; morula, $n = 2$; blastocyst, $n = 3$; 6-week embryos, $n = 2$). P values are shown (one-sided Wilcoxon rank-sum test). **e**, Reduced CpG methylation level (ML) ($\text{ML}_{\text{sperm}} - \text{ML}_{\text{blastocyst}}$) in A compartments and B compartments of human blastocyst (biological replicates pooled; $n = 3$). P value is calculated by one-sided Wilcoxon rank-sum test. **f**, Reduced methylation level ($\text{ML}_{\text{sperm}} - \text{ML}_{\text{blastocyst}}$) for regions with A-to-A, A-to-B, B-to-A, and B-to-B compartment switches from human sperm to blastocysts (replicates pooled; $n = 3$). *** $P < 0.001$ for comparisons between 'A-to-A' and other groups (two-sided Wilcoxon rank-sum test). Right, randomly shuffled A/B compartment switched regions as control.



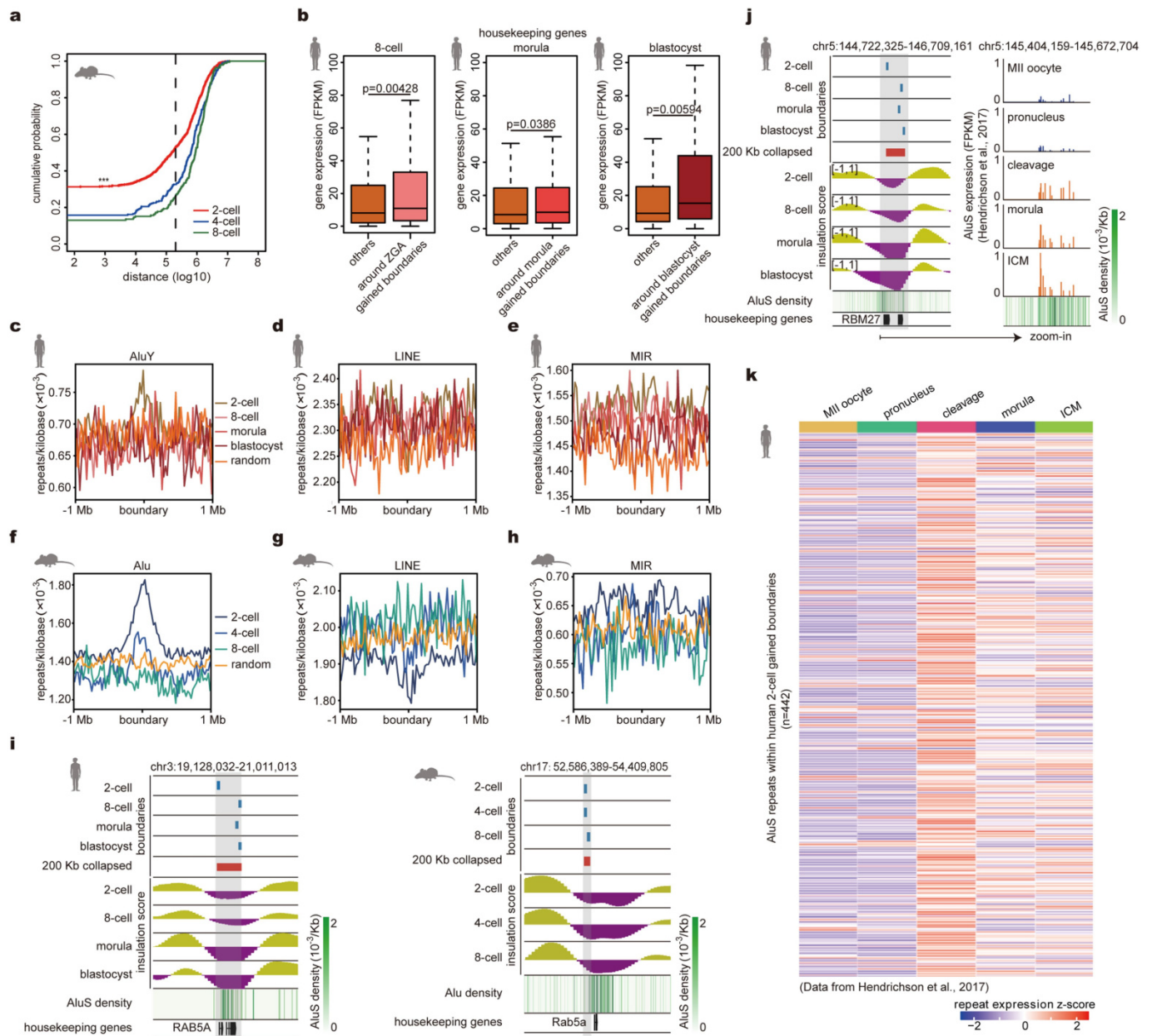
Extended Data Fig. 5 | Human sperm have no typical TADs. **a**, Interaction heat maps around *Hoxa* cluster in mouse sperm^{8,16} and mouse E3.5 embryos⁸. **b**, Density plot of interaction insert size for human sperm, human blastocysts and mouse sperm (Methods). **c**, Contact probability decay curve for human sperm, mouse sperm and metaphase HeLa cells. **d, e**, Examples of interaction heat maps at human and mouse conserved syntenic regions from unmixed sperm and mixed human and mouse sperm sample. Interaction heat maps in **d** (right) and **e** (right) are for HeLa and HT22 cells somatic cell mixed samples, respectively. **f**, Pearson correlation heat maps of chromosome 2 at 500-kb resolution in human sperm. Left, unmixed human sperm Hi-C. Right, human

sperm from the mixed sperm sample (replicates pooled). **g**, RAD21 western blots and Ponceau staining of samples from somatic cell lines and human and mouse sperm. Black arrow indicates RAD21 band. Experiment repeated on two biologically independent replicates per sample. For gel source data, see Supplementary Fig. 1. **h**, Right, CTCF western blots in the control and knockdown cell lines (control and siCTCF HEK 293; control and siCTCF HeLa). Left, Ponceau staining. Black arrow indicates CTCF band. Experiment repeated on two biologically independent replicates per sample. For gel source data, see Supplementary Fig. 1.



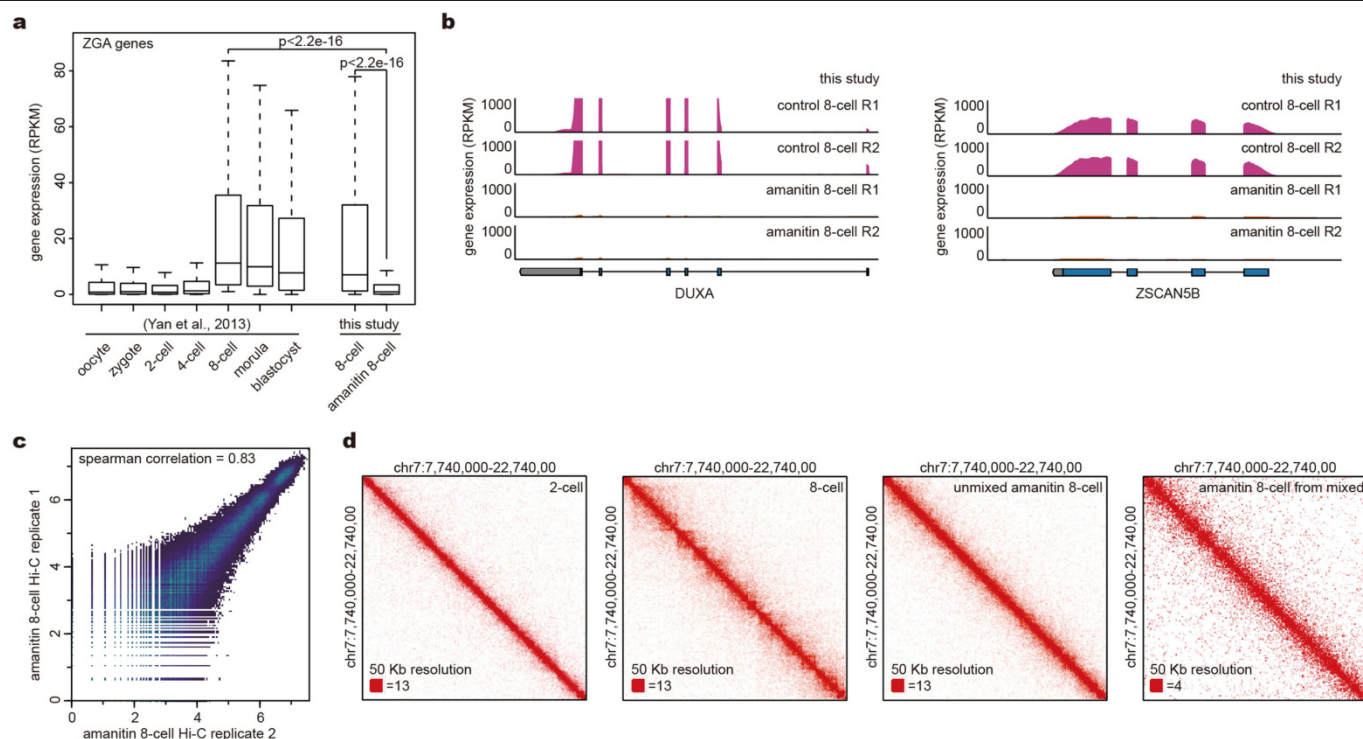
Extended Data Fig. 6 | Insulated boundary dynamics in human embryos and mouse embryos. a, An example for insulated boundaries in human embryos as yellow arrows point in the 2-cell stage. Dark coloured regions in TAD separation score heat maps mean strong insulation. The bottom blue bars are insulated boundaries of each stage. The red bars are 200-kb collapsed boundaries. Grey

boxes highlight the 200-kb collapsed boundaries. **b,** Heat map illustrating the insulation score at stage-specific gained insulated boundaries in mouse embryos. **c, d,** An example of shared ZGA boundaries in human embryos (**c**) and mouse embryos (**d**). Grey boxes highlight the shared 200-kb collapsed ZGA boundaries.



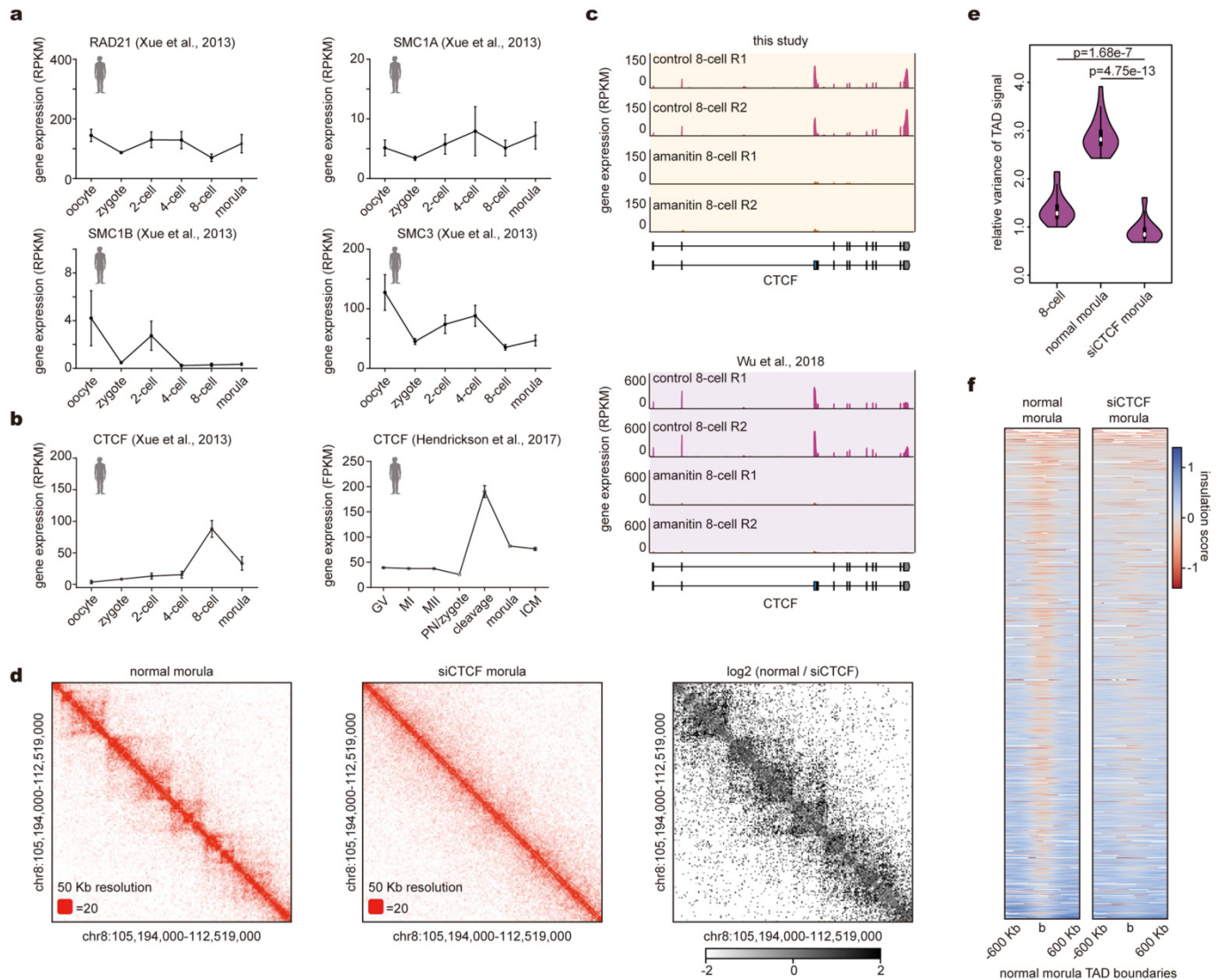
Extended Data Fig. 7 | Analysis of housekeeping genes and repeat elements at insulated boundaries. **a**, Cumulative distribution function plot for the distance of stage-specific gained insulated boundaries to the closest housekeeping genes in mouse embryos⁸ (pooled data from 2–3 biological replicates). The dash line marks a distance of 200 kb; 2-cell versus 4-cell, $P=1.25 \times 10^{-9}$; 2-cell versus 8-cell, $P=1.59 \times 10^{-11}$ (two-sided Kolmogorov–Smirnov test). **b**, Gene expression (expression data from ref.²²; 8-cell, $n=20$; morula, $n=16$; blastocyst, $n=30$) for housekeeping genes within 200 kb of the stage-specific gained boundaries and the other housekeeping genes at the stages when insulated boundaries are being formed. Boxes represent the 25th, 50th and 75th percentiles and whiskers show $1.5 \times$ the interquartile range. P values are shown (two-sided Wilcoxon rank-sum test). **c–e**, Enrichment analysis of repeat elements at stage-specific gained insulated boundaries in human embryos.

Enrichment analysis of AluY repeats (**c**), LINE repeats (**d**) and MIR repeats (**e**). **f–h**, Enrichment analysis of repeat elements at stage-specific gained insulated boundaries in mouse embryos. Enrichment analysis of Alu repeats (**f**), LINE repeats (**g**) and MIR repeats (**h**). **i**, Examples of the earlier stage gained insulated boundary locating around Alu dense regions in human and mouse. **j**, Left, snapshot of insulated boundaries overlaid with insulation score, AluS density and housekeeping genes. Grey box highlights the 200-kb collapsed boundary. Right, expression of AluS repeats at this boundary. **k**, Expression z-score of AluS repeats locating at human 2-cell gained boundaries during human embryonic development (repeat expression data from ref.²¹; $n=2$ for each stage; for repeat expression calculation, see Methods). Average gene expression of AluS repeats within a boundary was used. The number of 2-cell gained boundaries with AluS repeats is shown ($n=442$).



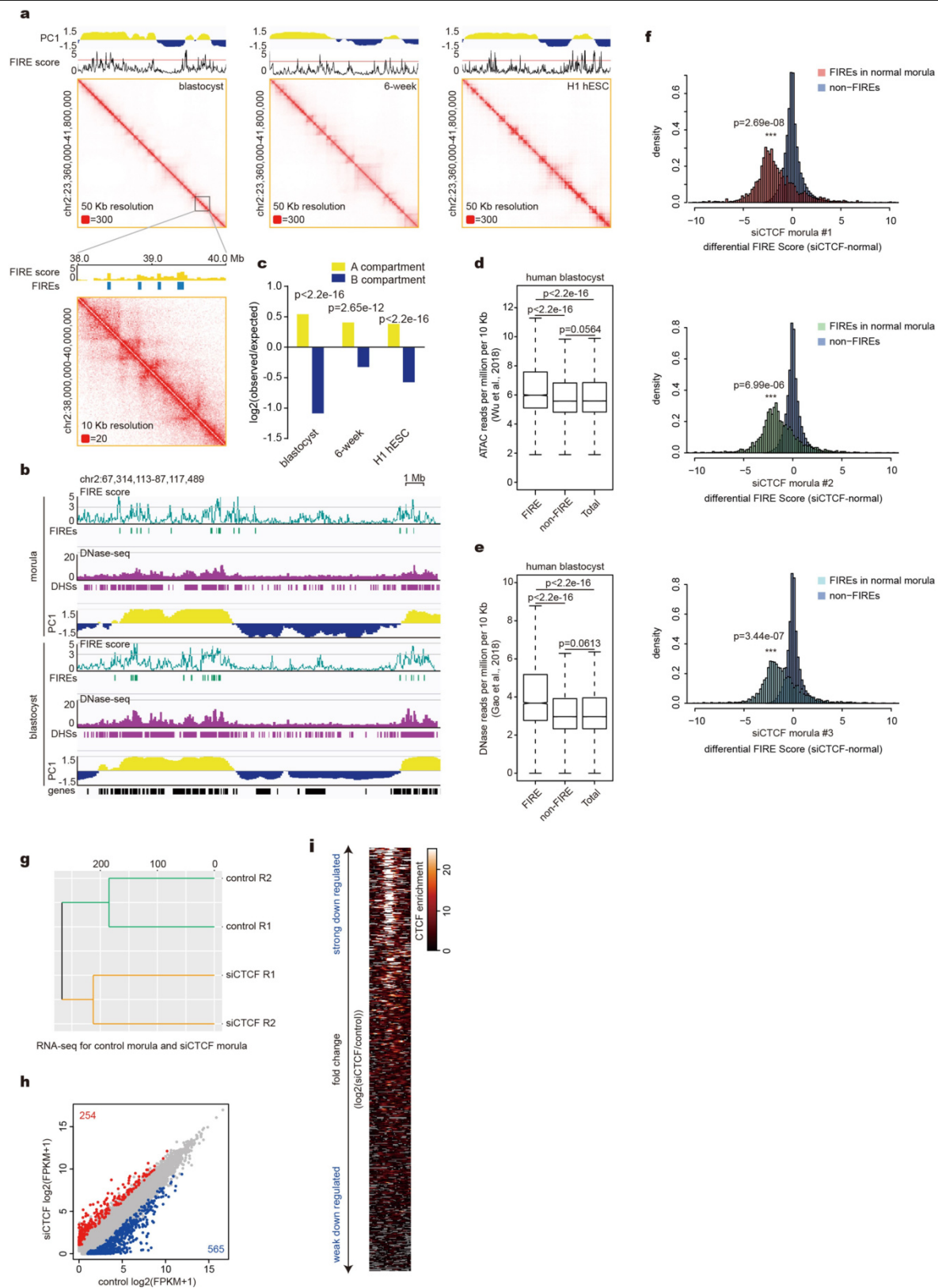
Extended Data Fig. 8 | TAD establishment depends on ZGA in human embryos. **a**, ZGA gene expression in human embryos and α -amanitin treated 8-cell embryos (pooled replicates; for data from ref.²²: oocyte, $n = 3$; zygote, $n = 3$; 2-cell, $n = 6$; 4-cell, $n = 12$; 8-cell, $n = 20$; morula, $n = 16$; blastocyst, $n = 30$; for data from this study: 8-cell, $n = 2$; α -amanitin treated 8-cell, $n = 2$). ZGA genes were identified as in ref.²² (Methods). P values are calculated by the one-sided

Wilcoxon rank-sum test. Boxes represent the 25th, 50th and 75th percentiles and whiskers show $1.5 \times$ the interquartile range. **b**, Gene expression for human ZGA genes *DUXA* and *ZSCAN5B* in control 8-cell and α -amanitin treated 8-cell embryos. **c**, The Spearman correlation for two α -amanitin treated 8-cell Hi-C replicates. **d**, Interaction heat maps of human 2-cell, 8-cell, unmixed α -amanitin treated 8-cell embryos and α -amanitin 8-cell from the mixed sample (Methods).



Extended Data Fig. 9 | CTCF regulates TAD establishment in human embryos. **a**, Gene expression for different cohesin complex subunits during human embryonic development. Expression data from ref. ²⁰: oocyte, $n = 3$; zygote, $n = 2$; 2-cell, $n = 3$; 4-cell, $n = 4$; 8-cell, $n = 11$; morula, $n = 3$. Data are mean \pm s.e.m. **b**, Dynamics of *CTCF* expression during human embryonic development. Expression data from refs. ^{20,21}. Data are mean \pm s.e.m. **c**, *CTCF* gene expression in control 8-cell and α -amanitin treated 8-cell embryos. **d**, Interaction heat maps of untreated control morula and *siCTCF* morula with equal reads at 50-kb resolution. Right, contrast heat map between control and

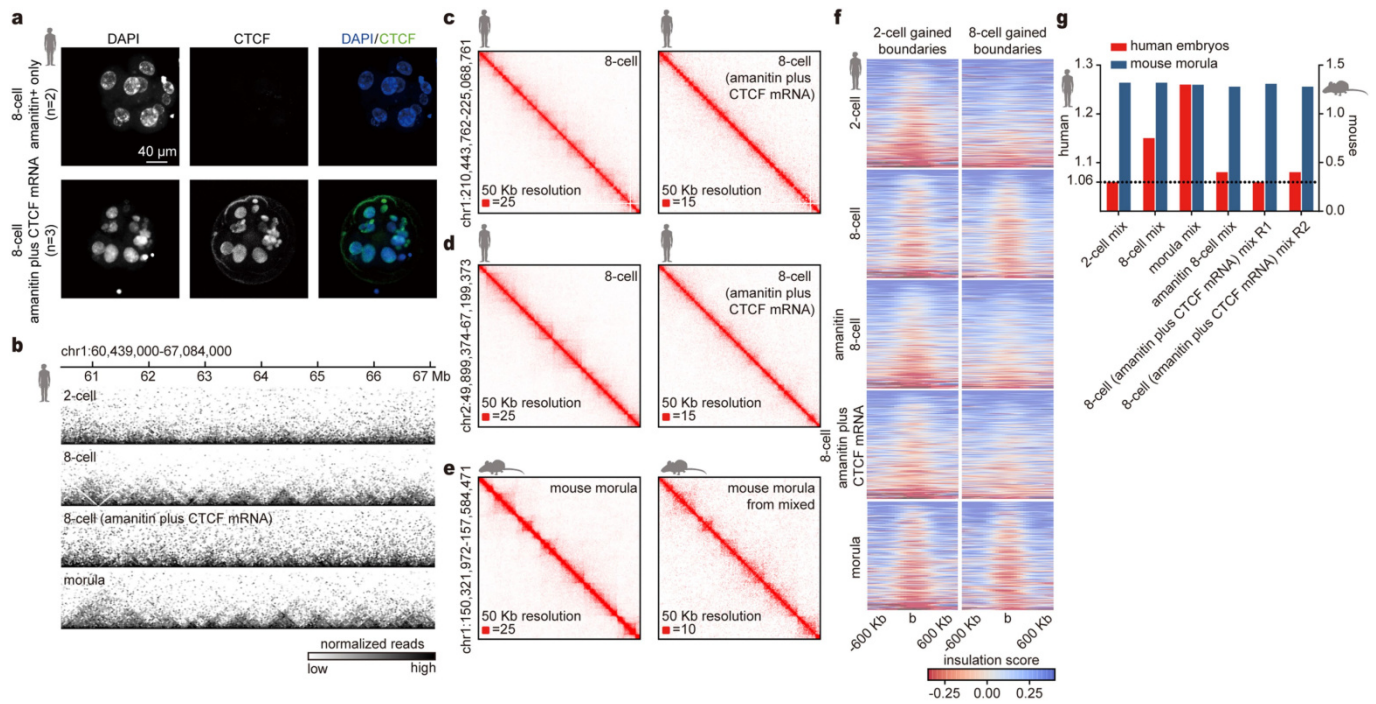
siCTCF morula. Dark colour represents increased interactions in untreated morula compared with *siCTCF* morula. **e**, Violin plot for relative variance of TAD signal in 8-cell, control morula and *siCTCF* morula (equal number of reads generated from 2–3 biological replicates for each stage). *P* values are shown (two-sided Wilcoxon rank-sum test). The white boxes in the violin plot represent median values. **f**, Enrichment heat maps for the insulation score of control morula and *siCTCF* morula at control morula TAD boundaries (± 600 kb). b, TAD boundary centre.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | The change of FIRE score and gene expression in siCTCF morula. **a**, Snapshot of FIREs in human blastocysts, 6-week embryos and H1 human ES cells overlaid with PC1 tracks and FIRE score tracks. One zoomed-in region in human blastocysts is also shown. **b**, Snapshot of FIREs in human morula and blastocysts overlaid with DHSs²⁵ and A/B compartments. **c**, Bar plot showing FIREs enrichment and depletion in A/B compartments (replicates pooled; blastocyst, $n = 3$; 6-week, $n = 2$; H1 human ES cell, $n = 2$). P values are also shown (χ^2 test). **d**, Box plot showing ATAC-seq reads²⁴ signal at FIREs, non-FIREs and whole genome in human blastocysts. Boxes represent the 25th, 50th and 75th percentiles and whiskers show 1.5× the interquartile range. P values are also shown (two-sided Wilcoxon rank-sum test). **e**, Box plot showing DNase-seq reads²⁵ signal at FIREs, non-FIREs and whole genome in human blastocysts. Boxes represent the 25th, 50th and 75th percentiles and

whiskers show 1.5× the interquartile range. P values are also shown (two-sided Wilcoxon rank-sum test). **f**, Histograms of FIRE score difference between siCTCF and untreated control morula (siCTCF – untreated) in FIREs and non-FIREs for siCTCF morula no. 1, siCTCF morula no. 2 and siCTCF morula no. 3. P values for differential FIREscore (siCTCF – untreated) in FIREs and non-FIREs are also shown (one-sided t -test). **g**, The hierarchical cluster of gene expression in control morula and siCTCF morula analysed using the package gg dendro. **h**, Scatter plot of gene expression between control morula and siCTCF morula. Red dots refer to upregulated genes (254 genes) in siCTCF morula. Blue dots refers to downregulated genes (565 genes) in siCTCF morula. **i**, Heat map for human ES cell CTCF ChIP-seq signal⁴⁷ around gene TSSs downregulated in siCTCF morula. Strongly downregulated genes ($\log_2(\text{fold change}(\text{siCTCF}/\text{control})) < -5$).



Extended Data Fig. 11 | TADs cannot re-establish in 8-cell (α -amanitin plus *CTCF* mRNA) embryos. **a**, Immunofluorescence confocal images of CTCF in α -amanitin treated 8-cell embryos ($n=2$) and 8-cell (α -amanitin plus *CTCF* mRNA) embryos ($n=3$). Scale bar, 40 μ m. **b**, Track snapshot for TAD structures in 8-cell (α -amanitin plus *CTCF* mRNA) embryos along with untreated 2-cell, 8-cell and morula embryos. **c, d**, Interaction heat map examples at 50-kb resolution for human 8-cell embryos and 8-cell (α -amanitin plus *CTCF* mRNA) embryos. **e**, Interaction heat map examples for mouse morula embryos without

mix and mouse morula from the α -amanitin plus *CTCF* mRNA mixed sample (Methods). **f**, Enrichment heat maps for the insulation score of human 2-cell embryos, 8-cell embryos, α -amanitin 8-cell embryos, 8-cell (α -amanitin plus *CTCF* mRNA) embryos and morula embryos around boundaries (± 600 kb). **g**, Bar plots for TAD structure strength of human embryos and mouse morula from mixed samples. Left y axis, human embryos; right y axis, mouse morula.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection ZEN software (2010) was used to collect confocal microscopy data

Data analysis HiCEXplorer suite (version 2.1);juicer tool;GenomeDISCO software;'matrix2insulation.pl' <https://github.com/dekkerlab/giorgett-nature-2016>;HOMER;'pairsqc.py' (<https://github.com/4dn-dcic/pairsqc>);UCSC liftOver tool online;DeepTools2 suite;R;Prism;ZEN software (2010);ImageJ 1.52a; See details in Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

CRA000852 and CRA000108

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The number of embryos used in this study is shown in Supplementary Table 1 and Supplementary Table 6. About as few as 50-100 cells can be used for a Hi-C assay, and 1 embryos can be used to prepare a RNA-seq library. In our analysis, our ultra-low-input Hi-C using 50-100 cells and SMART-seq2 RNA-seq can provide robust results (Extended Data Figure 1). No statistical method were used to predetermine the sample size.
Data exclusions	No data were excluded from the analyses.
Replication	Hi-C were performed at least 3 biological replicates for each stage of preimplantation embryos. And RNA-seq were performed at least 2 biological replicates for each sample. Detailed information is shown in Supplementary Table 1 and Supplementary Table 6. All replications were consistent for data results.
Randomization	All the embryos were collected and randomly allocated for experiments.
Blinding	During the siRNA injection , investigators injected the siRNA to the embryos without knowing the types of the siRNA. For other experiments, variations were controlled through replicates.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.</i>
-------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.
Did the study involve field work?	<input type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access and import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<p>Antibodies used in western blotting:</p> <p>Anti-CTCF antibody, abcam, cat#: ab188408, lot#GR3253930-3, 1:1000 dilution</p> <p>Anti-RAD21 antibody, abcam, cat#: ab992, lot#GR239192-1, 1:1000 dilution</p> <p>Goat Anti-Rabbit IgG, CWBIO, Cat no: CW0103, lot#013341/50402, 1:3000 dilution</p> <p>Immunostaining Antibodies:</p> <p>Anti-CTCF antibody, abcam, cat#: ab188408, lot#GR3253930-3, 1:500 dilution</p> <p>Alexa Fluor 488-labeled goat anti-rabbit IgG, Beyotime, cat#: P0176, 1:1000 dilution</p>
Validation	All the antibodies used in this study were commercial antibodies and were only used for applications, with validation procedures

described on the following sites of the manufacturers: <https://www.abcam.com>; <https://www.beyotime.com>; <https://www.cwbiotech.com>.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK 293 cells (ATCC: ATCC® CRL-1573™); HeLa (ATCC: ATCC® CCL-2™); 4T1 (ATCC: ATCC® CRL-2539™); HT22 (EMD Millipore, SCC129)
Authentication	All these cell lines were used in several papers published by our lab, no further authentication were performed for these cell lines.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Palaeontology

Specimen provenance	<i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).</i>
Specimen deposition	<i>Indicate where the specimens have been deposited to permit free access by other researchers.</i>
Dating methods	<i>If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.</i>

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Mouse sperm cells were collected from the cauda epididymis in 8 weeks old C57BL/6J male mice (Vital River). The mouse morula embryos were from the cross of C57 BL/6J female (Vital River; 4-6 weeks) and PWK/PhJ male (Jackson Laboratory; 10 weeks in average).
Wild animals	No wild animals were used.
Field-collected samples	No field collected samples
Ethics oversight	All animal maintenance and experimental procedures were carried out according to guidelines of Institutional Animal Care and Use Committee (IACUC) of Beijing Institute of Genomics, CAS, Beijing, China.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The oocyte donors are 25-38 year-old women with normal BMI. The healthy sperm donors are 22-50 year-old men with normal semen. All volunteers have no family heredity case history, contagion case and smoke history. The gametes and embryos are checked under microscopy and are with high quality. The donated frozen embryos as well as the donated oocytes and sperm in this study are from Center for Reproductive Medicine, Shandong University.
Recruitment	Research donors were recruited from patients taking in vitro fertilization treatments. Before giving consent, donors have a suitable opportunity to receive proper counselling about the implications of the donation and potential risks. Written informed consent was obtained from all oocyte and sperm donors, respectively. They were informed that the donation would not affect their IVF process.
Ethics oversight	Institutional Review Board of Reproductive Medicine, Shandong University and Beijing Institute of Genomics

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.
Study protocol	Note where the full trial protocol can be accessed OR if not available, explain why.
Data collection	Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.
Outcomes	Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.
Files in database submission	Provide a list of all files available in the database submission.
Genome browser session (e.g. UCSC)	Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.
Instrument	Identify the instrument used for data collection, specifying make and model number.
Software	Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance	Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.
Gating strategy	Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	Indicate task or resting state; event-related or block design.
Design specifications	Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.
Behavioral performance measures	State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)	Specify: functional, structural, diffusion, perfusion.	
Field strength	Specify in Tesla	
Sequence & imaging parameters	Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.	
Area of acquisition	State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.	
Diffusion MRI	<input type="checkbox"/> Used	<input type="checkbox"/> Not used

Preprocessing

Preprocessing software	Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).
Normalization	If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.
Normalization template	Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.
Noise and artifact removal	Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).
Volume censoring	Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings	Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).
Effect(s) tested	Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.
Specify type of analysis:	<input type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See Eklund et al. 2016)	Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.
Correction	Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

- n/a | Involved in the study
- ☐ ☐ Functional and/or effective connectivity
- ☐ ☐ Graph analysis
- ☐ ☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.

Light-driven anaerobic microbial oxidation of manganese

<https://doi.org/10.1038/s41586-019-1804-0>

Received: 8 June 2018

Accepted: 8 October 2019

Published online: 4 December 2019

Mirna Daye^{1*}, Vanja Klepac-Ceraj², Mihkel Pajusalu¹, Sophie Rowland², Anna Farrell-Sherman², Nicolas Beukes³, Nobumichi Tamura⁴, Gregory Fournier¹ & Tanja Bosak^{1*}

Oxygenic photosynthesis supplies organic carbon to the modern biosphere, but it is uncertain when this metabolism originated. It has previously been proposed^{1,2} that photosynthetic reaction centres capable of splitting water arose by about 3 billion years ago on the basis of the inferred presence of manganese oxides in Archaean sedimentary rocks. However, this assumes that manganese oxides can be produced only in the presence of molecular oxygen³, reactive oxygen species^{4,5} or by high-potential photosynthetic reaction centres^{6,7}. Here we show that communities of anoxygenic photosynthetic microorganisms biomineralize manganese oxides in the absence of molecular oxygen and high-potential photosynthetic reaction centres. Microbial oxidation of Mn(II) under strictly anaerobic conditions during the Archaean eon would have produced geochemical signals identical to those used to date the evolution of oxygenic photosynthesis before the Great Oxidation Event^{1,2}. This light-dependent process may also produce manganese oxides in the photic zones of modern anoxic water bodies and sediments.

Manganese and more than 30 of its described oxides and hydroxides mediate the cycling of various trace metals and nutrients in the environment. The ability of microorganisms to oxidize Mn(II) anaerobically has also been hypothesized to have been a critical step in the evolution of oxygenic photosynthesis on the early Earth⁶. However, modern microorganisms are not known to anaerobically oxidize manganese. Here we demonstrate this activity in active microbial cultures that grow in the presence of nanomolar concentrations of oxygen, relevant for the Archaean Earth.

Inoculum for the enrichment cultures of strictly anaerobic, photosynthetic biofilms came from the meromictic Fayetteville Green Lake (New York, USA). The most abundant phototroph in the sulfidic photic zone of this lake is the green sulfur bacterium *Chlorobium* sp.⁸, a microorganism that uses sulfide as the electron donor for photosynthesis. Biofilms (Fig. 1a) containing *Chlorobium* sp. and other strict anaerobic microorganisms were enriched in a minimal medium amended with 20–50 μ M Na₂S and 1 mM MnCl₂ and equilibrated with an anaerobic atmosphere of 80% N₂ and 20% CO₂ at pH 7. The concentration of O₂ in the medium was lower than 2 nM during the entire experiment, both in the presence and the absence of cells (Extended Data Fig. 1, Methods). These experimental concentrations match the upper estimates for the Archaean Earth⁹. The anaerobic medium also lacked other potential oxidants for Mn(II), such as nitrite, nitrate and H₂O₂, and these species were not produced in sterile controls ('Concentrations of dissolved species in culture media' in the Methods).

Four times more biomass grew in photosynthesizing cultures relative to the cultures incubated in the dark ($P < 0.001$) (Fig. 1b). The enrichment protocol yielded a stable community that contained *Chlorobium*, *Paludibacter*, *Acholeplasma*, *Geobacter*, *Desulfomicrobium*,

Clostridium, *Acetobacterium* and several other bacteria (Fig. 1c). The *Chlorobium* sp. was the most abundant taxon across all conditions, as well as the only identifiable phototroph (Fig. 1c). Its genome was 98.4% similar to *Chlorobium limicola* Frassasi.

This microbial community was essential for the precipitation of minerals and oxidation of manganese. Manganese-rich dolomite was the most abundant precipitate in photosynthetic cultures amended with 0.1–1 mM Mn(II) and 20–250 μ M sulfide¹⁰ (Fig. 2, Extended Data Fig. 2a, b, d), which is comparable to the estimated dissolved Mn(II) concentrations (0.1 mM) maintained by the precipitation of manganese-containing carbonate minerals on Archaean carbonate platforms^{11,12}. Cultures incubated in this range of chemical conditions also contained manganese oxide minerals (Figs. 1d, 2), but cultures incubated with approximately 1 μ M Mn(II) or 1 mM sulfide did not (Extended Data Fig. 2c). Precipitates were entirely absent from sterile controls incubated in the light, whereas minor calcite, less dolomite and no manganese oxides were detected in cultures shielded from the light (Fig. 2). Elemental sulfur (S⁰) accumulated in cultures that did not contain oxidized manganese minerals (Extended Data Fig. 3c), as expected during the growth of *Chlorobium* sp. under conditions in which sulfide was the main photosynthetic electron donor. Abiotic reactions did not contribute detectable amounts of manganese oxides or other minerals under our experimental conditions. Instead, the microbial presence and photosynthetic activity strongly controlled the nucleation and precipitation of minerals, including manganese oxides.

To characterize the redox cycling of manganese, we examined its oxidation state by surface-sensitive methods (Extended Data Fig. 3). After two weeks, sterile controls and cultures that were incubated in the dark contained only Mn(II) (Extended Data Fig. 3a). Dolomite that formed

¹Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Department of Biological Sciences, Wellesley College, Wellesley, MA, USA. ³DST-NRF CIMERA, Geology Department, University of Johannesburg, Johannesburg, South Africa. ⁴Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. *e-mail: mirnadaye@gmail.com; tbosak@mit.edu

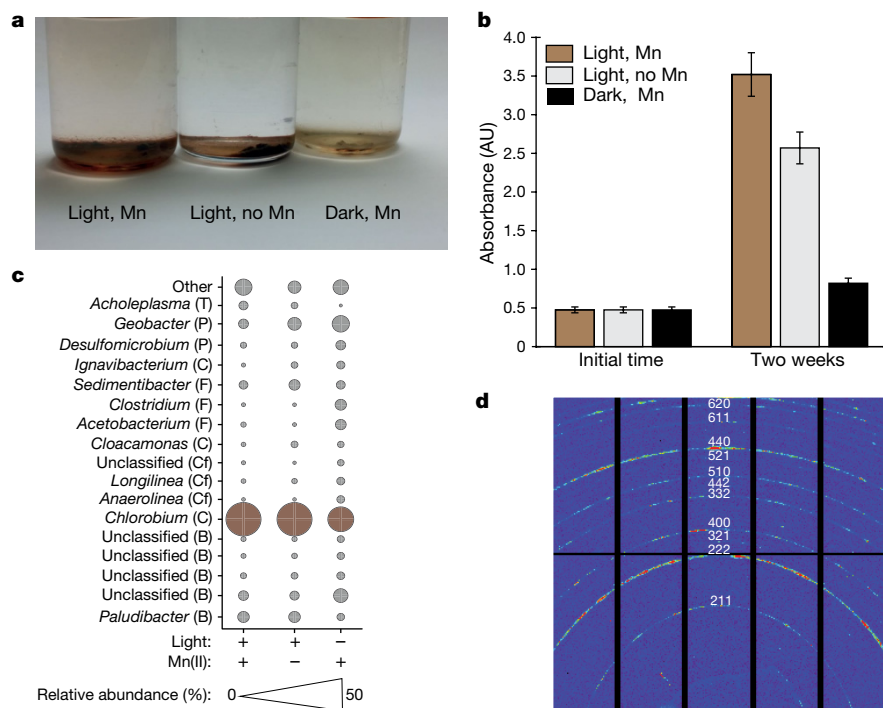


Fig. 1 | Biofilms incubated for two weeks. **a**, Dark brown biofilm incubated in the light with Mn(II) (left), less-extensive biofilm incubated in the light without Mn(II) (middle) and a yellow biofilm incubated in the dark with Mn(II) (right). **b**, Biomass of biofilms measured by the crystal violet assay. All incubations were performed in triplicate. AU, arbitrary units. $n = 3$ measurements; error bars ± 0.3 s.d. **c**, Microbial diversity in biofilms obtained by high-throughput

Illumina sequencing of 16S rRNA genes. The most abundant taxon across all conditions was a *Chlorobium* sp. (30–60%). This microorganism was more abundant in photosynthesizing cultures. B, Bacteroidetes; C, Chlorobi; Cf, Chloroflexi; F, Firmicutes; P, Proteobacteria; T, Tenericutes. **d**, Diffraction pattern indices for Mn_2O_3 in the spectrum acquired by synchrotron micro-focused X-ray diffraction (μXRD) of a biofilm incubated in the light for two weeks.

in photosynthetic cultures contained Mn(II)¹⁰, but we also detected Mn(II), Mn(III) and Mn(IV) in calcium manganese oxides, Mn_2O_3 and other minerals (Figs. 1d, 2, 3, Extended Data Fig. 3). The presence of manganese oxides in sulfidic photosynthetic cultures was surprising,

but they were repeatedly found in biofilms that were between two weeks and two months old (Fig. 2). A colorimetric assay quantified $5.1 \pm 0.8 \mu\text{M}$ of oxidized manganese in one-week-old biofilms, $7.2 \pm 0.8 \mu\text{M}$ oxidizing equivalents in two-week-old biofilms and $>10 \mu\text{M}$ in three-week

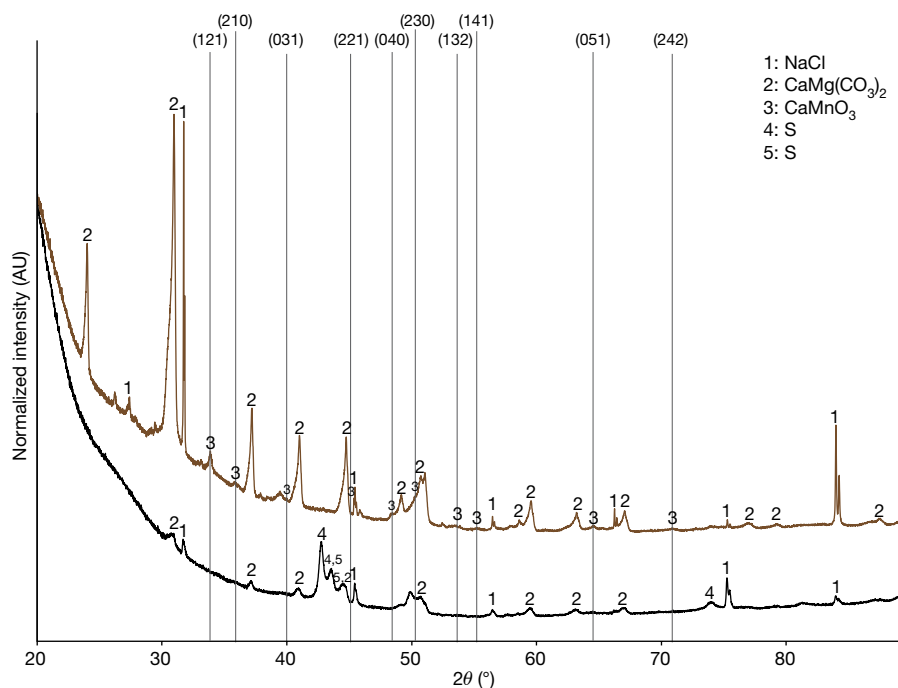


Fig. 2 | XRD spectra of biofilms incubated for two weeks. Top spectrum, biofilms incubated in the light with 1 mM Mn(II). Black lines indicate Miller indices (hkl) assigned to each peak of CaMnO_3 . Bottom spectrum, biofilms incubated in the dark with 1 mM Mn(II). The sulfur phases (S) formed during the

treatment to remove oxidized manganese minerals before inoculation ('Modified FGL medium' and 'Interpretation of XRD peaks' in the Methods). All data are representative of two independent experiments.

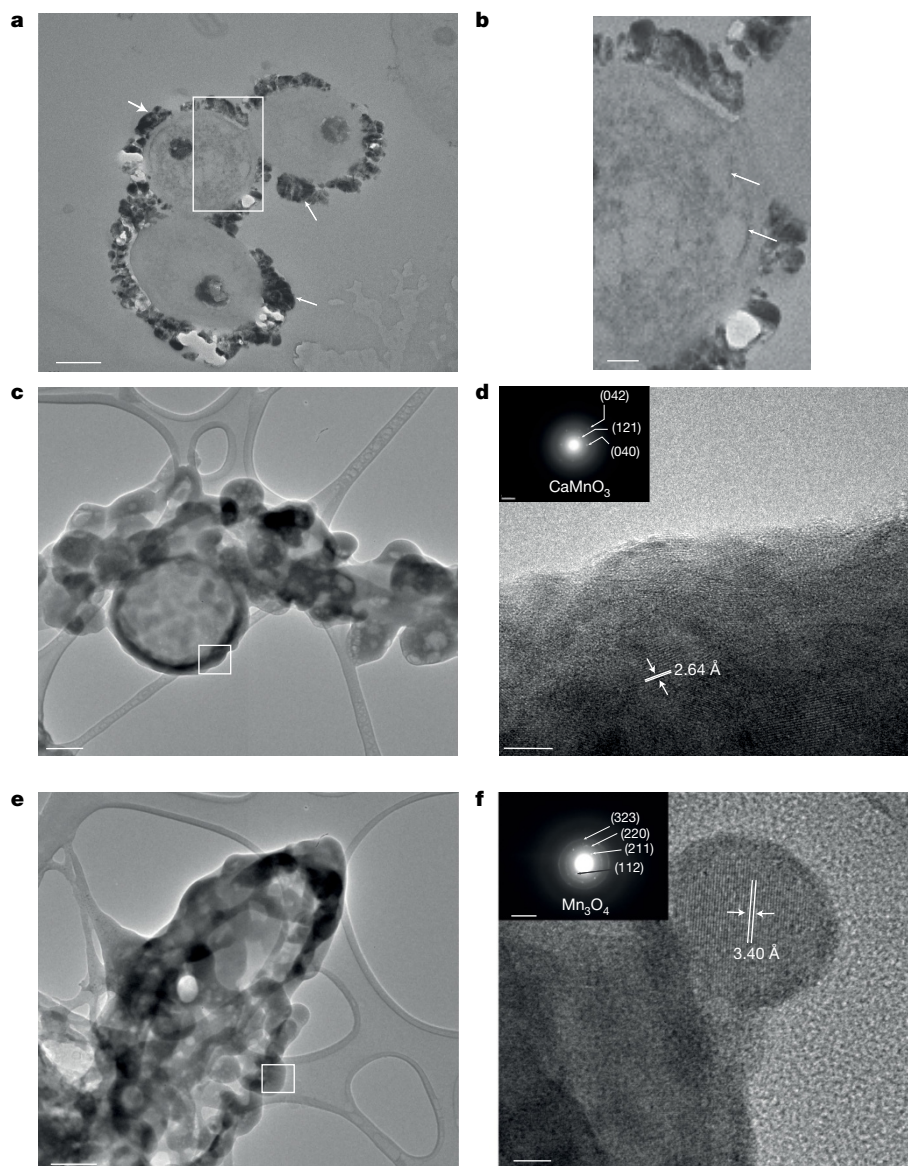


Fig. 3 | Interactions between minerals and microorganisms in biofilms incubated with Mn(II) in the light for two weeks. a, Mineral-encrusted cells (white arrows) in fixed and stained samples (transmission electron microscopy (TEM) at 80 kV); the white square indicates the area magnified in **b**. Scale bar, 200 nm. **b,** Chlorosomes (white arrows) in a cell encrusted by manganese oxide precipitates (TEM at 80 kV, fixed and stained sample). Scale bar, 50 nm. **c,** TEM at 200 kV of unprocessed and unstained mineral-encrusted microbial cells. White square indicates the region selected for selected area electron diffraction (SAED), shown in **d**. Scale bar, 200 nm. **d,** High-resolution TEM and

SAED of minerals around an unstained cell from a fresh suspension of the microbial culture on a TEM grid show CaMnO_3 with a d -spacing of 2.64 Å. Scale bars, 5 nm (**d**), 51/nm (inset). **e,** TEM at 200 kV of unprocessed and unstained encrusted microbial cells. White square indicates the region selected for SAED shown in **f**. Scale bar, 200 nm. **f,** High-resolution TEM and SAED of minerals in the area outlined by the white square in **e** show Mn_3O_4 with a d -spacing of 3.40 Å. Scale bars, 5 nm (**f**), 101/nm (inset). All data are representative of two independent experiments.

old biofilms ('Quantification of biofilms and oxidized manganese in biofilms' in the Methods). All oxidized manganese was determined as KMnO_4 equivalents and none was detected in dark controls.

Oxidized manganese was present in minerals that were only found on cell surfaces (Fig. 3a) or around extracellular vesicles. These cells could be identified as *Chlorobium* sp. on the basis of the presence of large intracellular complexes of photosynthetic antennae called chlorosomes (Fig. 3b) and surface protrusions called spinae¹³. High-resolution transmission electron micrographs of fresh cell suspensions showed manganese–calcium minerals with a uniform lattice fringe that corresponded to the (121) plane with interplanar spacing of 2.64 Å of calcium manganese oxide (Fig. 3d), Mn_3O_4 (Fig. 3f) and other manganese minerals (Extended Data Fig. 2). Extracellular vesicles, spinae and

manganese oxide minerals were absent from *Chlorobium* sp. when the cultures were incubated with Mn(II) in the dark or photosynthetically with 1 mM sulfide.

Light-driven manganese oxidation occurred only when *Chlorobium* sp. and other microorganisms, including *Geobacter* sp., were growing together. Oxidized manganese was present in enrichment cultures of microorganisms from Fayetteville Green Lake that contained *Chlorobium* sp., *Geobacter* sp., *Acholeplasma equifetale*, *Alistipes* sp. HGB5 and *Caldicoprobacter oshimai*, but absent from co-cultures of *Chlorobium* sp. and *Desulfomicrobium* sp. (Extended Data Fig. 4) and pure cultures of *C. limicola*. Manganese oxidizing activity was also detected in the co-cultures containing *C. limicola*, *Chlorobaculum tepidum* and *Geobacter lovleyi* (DSM 245, DSM 12025 and DSM 17278, DSMZ) ('Probing

the redox state of manganese in co-cultures' and 'Interpretation of XPS spectra' in the Methods) and the co-cultures of *C. limicola* and *G. lovleyi* (Extended Data Fig. 5). This suggests that the activity may depend on extracellular electron transfer between the latter two organisms¹⁴ by a currently unknown mechanism. The only photosynthetic reaction centre that the genome of *Chlorobium* sp. encodes is a well-studied centre with a midpoint potential around +250 mV¹⁵ and that cannot directly oxidize Mn(II) bicarbonate ($E_h = 520\text{--}670$ mV)^{6,16}. *Chlorobium* sp. also lacks clear homologues of proteins that are known to oxidize manganese under aerobic conditions^{17–19}. Manganese oxidation in *C. limicola* may occur by an endergonic mechanism, analogous to that proposed for the oxidation of nitrite by the *Thiocapsa* sp. strain KS1²⁰. The electron transfer between *C. limicola* and *G. lovleyi* may involve high-potential cytochrome *c* in *Chlorobium* sp. and OmpB operating in reverse in *Geobacter* sp., but the Mn(II) oxidation mechanisms and potential oxidants in the co-cultures of *C. limicola* and *G. lovleyi* remain to be characterized.

The abundance of oxidized manganese in microbial biofilms, and its absence from the sterile controls, shows that microbial consortia can mediate the precipitation of manganese oxide minerals under conditions similar to those in the Archaean eon. These findings expand the diversity of minerals and redox processes beyond what was thought possible in strictly anaerobic environments or in the presence of high-potential photosynthetic reaction centres. Microbial interactions that mediate the light-dependent redox cycling of manganese and couple it to other elemental cycles remain to be identified, but can be expected in modern environments in which light, sulfide and dissolved Mn(II) coexist and sulfide concentrations do not exceed about 0.2 mM (Extended Data Table 1). The light-dependent microbial production of manganese oxides is likely to stimulate the redox cycles of carbon, sulfur, nitrogen and iron, increase the diversity of anaerobic redox transformations (including nitrification–denitrification²¹) and influence interpretations of isotopic and chemical signatures of these processes in modern anaerobic settings.

The biological production of manganese oxides under Archaean-like chemical conditions has additional major implications for determining the timing of the origin of oxygenic photosynthesis, which is currently debated^{1,2,5,7}. The evolution of oxygenic photosynthesis¹ and the presence of locally oxic areas in Archaean marine systems prior to the Great Oxidation Event (GOE)²² are inferred from geochemical signals. However, interpretations of these signals assume the former presence of manganese oxides^{1,2,23,24}. For example, manganese carbonate deposits with the negative $\delta^{13}\text{C}$ values reported in the Neoproterozoic Sandur schist belt (India)²⁵ or Mesoproterozoic Witwatersrand–Mozaan strata (South Africa)^{22,23} are thought to have been produced by the microbial reduction of Mn(III) and Mn(IV) oxide minerals. In turn, these oxides are attributed to the aerobic oxidation of Mn(II) in the presence of oxygen. Molecular clock models are used to independently date the evolution of the crown group cyanobacteria and have produced estimates that range from before 3 billion years ago to after the GOE, depending on the sequence datasets used, prior assumptions made and specific model calibrations²⁶. These models support the radiation of anoxygenic green sulfur bacteria (such as *Chlorobium*) and green non-sulfur bacteria after the GOE²⁶, but also show that stem-group green sulfur bacteria diverged as early as 2.9 billion years ago. Given that the last common ancestor of modern green sulfur bacteria was photosynthetic, model estimates are consistent with anoxygenic photosynthesis within stem-group green sulfur bacteria long before the GOE (Extended Data Fig. 6). Therefore, anaerobic manganese oxidation that requires anoxygenic photosynthetic activity in the presence of sulfide could be as old as the anoxygenic phototrophic ancestors of any extant groups of phototrophs (including green sulfur bacteria, cyanobacteria or even an extinct lineage of anoxygenic phototrophs). Because any of these scenarios can pre-date the GOE, the relative contributions of anaerobic and oxygen-dependent Mn(II) oxidation to the redox texture

of sedimentary rocks much before the GOE²⁷ are an open question, and the loss of mass-independent sulfur isotope signals at the GOE²⁸ remains the firmest evidence for the biological production of oxygen.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1804-0>.

- Planavsky, N. J. et al. Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nat. Geosci.* **7**, 283–286 (2014).
- Crowe, S. A. et al. Atmospheric oxygenation three billion years ago. *Nature* **501**, 535–538 (2013).
- Tebo, B. M. et al. Biogenic manganese oxides: properties and mechanisms of formation. *Annu. Rev. Earth Planet. Sci.* **32**, 287–328 (2004).
- Oze, C., Sleep, N. H., Coleman, R. G. & Fendorf, S. Anoxic oxidation of chromium. *Geology* **44**, 543–546 (2016).
- Liang, M.-C., Hartman, H., Kopp, R. E., Kirschvink, J. L. & Yung, Y. L. Production of hydrogen peroxide in the atmosphere of a Snowball Earth and the origin of oxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* **103**, 18896–18899 (2006).
- Dismukes, G. C. et al. The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* **98**, 2170–2175 (2001).
- Johnson, J. E. et al. Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc. Natl Acad. Sci. USA* **110**, 11238–11243 (2013).
- Culver, D. A. & Brunskill, G. J. Fayetteville Green Lake, New York. V. Studies of primary production and zooplankton in a meromictic marl lake. *Limnol. Oceanogr.* **14**, 862–873 (1969).
- Pavlov, A. A. & Kasting, J. F. Mass-independent fractionation of sulfur isotopes in Archaean sediments: strong evidence for an anoxic Archaean atmosphere. *Astrobiology* **2**, 27–41 (2002).
- Daye, M., Higgins, J. & Bosak, T. Formation of ordered dolomite in anaerobic photosynthetic biofilms. *Geology* **47**, 509–512 (2019).
- Anbar, A. D. & Holland, H. D. The photochemistry of manganese and the origin of banded iron formations. *Geochim. Cosmochim. Acta* **56**, 2595–2603 (1992).
- Beukes, N. J. Facies relations, depositional environments and diagenesis in a major early Proterozoic stromatolitic carbonate platform to basinal sequence, Campbellrand Subgroup, Transvaal Supergroup, Southern Africa. *Sedim. Geol.* **54**, 1–46 (1987).
- Pibernet, I. V., IV & Abella, C. A. Sulfide pulsing as the controlling factor of spinae production in *Chlorobium limicola* strain UdG 6038. *Arch. Microbiol.* **165**, 272–278 (1996).
- Ha, P. T. et al. Syntrophic anaerobic photosynthesis via direct interspecies electron transfer. *Nat. Commun.* **8**, 13924 (2017).
- Prince, R. C. & Olson, J. M. Some thermodynamic and kinetic properties of the primary photochemical reactants in a complex from a green photosynthetic bacterium. *Biochim. Biophys. Acta* **423**, 357–362 (1976).
- Dasgupta, J., Tyryshkin, A. M., Kozlov, Y. N., Klimov, V. V. & Dismukes, G. C. Carbonate complexation of Mn²⁺ in the aqueous phase: redox behavior and ligand binding modes by electrochemistry and EPR spectroscopy. *J. Phys. Chem. B* **110**, 5099–5111 (2006).
- Butterfield, C. N., Soldatova, A. V., Lee, S.-W., Spiro, T. G. & Tebo, B. M. Mn(II,III) oxidation and MnO₂ mineralization by an expressed bacterial multicopper oxidase. *Proc. Natl Acad. Sci. USA* **110**, 11731–11735 (2013).
- Geszvain, K., Smesrud, L. & Tebo, B. M. Identification of a third Mn(II) oxidase enzyme in *Pseudomonas putida* GB-1. *Appl. Environ. Microbiol.* **82**, 3774–3782 (2016).
- Andeer, P. F., Learman, D. R., McIlvin, M., Dunn, J. A. & Hansel, C. M. Extracellular haem peroxidases mediate Mn(II) oxidation in a marine *Roseobacter* bacterium via superoxide production. *Environ. Microbiol.* **17**, 3925–3936 (2015).
- Hemp, J. et al. Genomics of a phototrophic nitrite oxidizer: insights into the evolution of photosynthesis and nitrification. *ISME J.* **10**, 2669–2678 (2016).
- Hulth, S., Aller, R. C. & Gilbert, F. Coupled anoxic nitrification/manganese reduction in marine sediments. *Geochim. Cosmochim. Acta* **63**, 49–66 (1999).
- Kurzweil, F., Wille, M., Ganter, N., Beukes, N. J. & Schoenberg, R. Manganese oxide shuttling in pre-GOE oceans—evidence from molybdenum and iron isotopes. *Earth Planet. Sci. Lett.* **452**, 69–78 (2016).
- Ossa Ossa, F., Hofmann, A., Vidal, O., Kramers, J. D. & Cavalazzi, G. Unusual manganese enrichment in the Mesoproterozoic Mozaan Group, Pongola Supergroup, South Africa. *Precamb. Res.* **281**, 414–433 (2016).
- Beukes, N. J., Gutzmer, J. & Nel, B. Ce anomalies in similar to 2.4 Ga iron and manganese formations as proxy for early oxygenation of oceanic environments. *Geochim. Cosmochim. Acta* **74**, A85 (2010).
- Manikyamba, C. & Naqvi, S. M. Geochemistry of Fe–Mn formations of the Archaean Sandur schist belt, India – mixing of clastic and chemical processes at a shallow shelf. *Precamb. Res.* **72**, 69–95 (1995).
- Magnabosco, C., Moore, K. R., Wolfe, J. M. & Fournier, G. P. Dating phototrophic microbial lineages with reticulate gene histories. *Geobiology* **16**, 179–189 (2018).
- Hazen, R. M. et al. Mineral evolution. *Am. Mineral.* **93**, 1693–1720 (2008).
- Bekker, A. et al. Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Culturing and sequencing

Enrichment and culturing conditions. Sediments were retrieved from Fayetteville Green Lake (FGL) by a metallic gravity scoop and stored at 4 °C in fully filled, hermetically sealed glass jars. These samples were used as inoculum for enrichment cultures in the FGL medium. All inoculations were conducted in an anaerobic chamber under a 5%CO₂/5%H₂/balN₂ (v/v/v) atmosphere using standard anaerobic techniques²⁹. In brief, FGL medium was flushed before and after autoclaving in hermetically sealed glass bottles. Sterile flushed FGL medium was inoculated inside the anaerobic chamber (5%CO₂/5%H₂/balN₂ (v/v/v) atmosphere). To avoid any issues associated with the exposure of biofilms to H₂, the serum bottles were opened inside the anaerobic chamber, inoculated in about 1 min, closed immediately, capped and flushed again with CO₂/N₂ for 60–75 min to remove H₂. All experiments were conducted in batch cultures and all cultures were inoculated by approximately 1 mg of biofilm that had been washed 6 times by anoxic nanopure water^{30,31}, mechanically dispersed by passing through a syringe and resuspended in sterile anaerobic medium. All these steps were carried out in the anaerobic glove box. All enrichment cultures were incubated at 27 °C with an incandescent white light bulb and a 12:12-h day:night cycle. All plasticware used in the anaerobic chamber was introduced into the chamber at least one week before the experiments.

The culture medium (FGL medium) contained 0.1 mM KH₂PO₄, 5.61 mM NH₄Cl, 0.9 mM KCl, 0.024 M NaHCO₃, 1 mM MnCl₂·2H₂O, 1 mM Na₂SO₄ and 1 ml/l of trace element solution. The trace element solution was prepared in 10% (v/v) HCl and contained, per litre, 1.5 g FeCl₂·4H₂O, 190 mg CoCl₂·6H₂O, 100 mg MnCl₂·4H₂O, 70 mg ZnCl₂, 31 mg Na₂MoO₄, 6 mg H₃BO₃, 2 mg CuCl₂·2H₂O. The pH of the medium was adjusted to 7 by the addition of NaOH (1 M) or HCl (1 M). After adjusting the pH to 7, the FGL medium was distributed into glass bottles of different volumes (12, 25, 50, 150 and 200 ml). The final background concentration of manganese was 0.4 µM. To inhibit the growth of oxygenic phototrophs, we added 0.01 mM 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) to the initial enrichments. All analyses described in the main text used later enrichments that were grown in DCMU-less medium. Glass serum bottles were capped by butyl rubber stoppers and aluminium seals. Before autoclaving, the FGL medium in bottles was flushed by 20% CO₂:80% N₂ for 1 h, the serum bottle headspaces for another 40 min total. The bottles were then autoclaved (40-min sterile). After autoclaving, the FGL medium in bottles were flushed again by 20% CO₂:80% N₂ for one hour, the serum bottle headspaces for another 40 min total after cooling.

A separately prepared selenium stock solution contained 2 mg of Na₂SeO₃ in 1,000 ml of 0.01 M NaOH. This solution was autoclaved and made anaerobic by flushing the bottles for 1 h and 40 min with 20% CO₂:80% N₂. The vitamin solution was prepared in nanopure water by aerobic filter sterilization and contained, per litre, 2 mg biotin, 2 mg folic acid, 10 mg pyridoxine-2H₂O, 5 mg thiamine-HCl-2H₂O, 5 mg riboflavin, 5 mg nicotinic acid, 5 mg D-Ca-pantothenate, 0.1 mg vitamin B12, 5 mg *p*-aminobenzoic acid and 5 mg lipoic acid.

The master stock solution (20×) contained 1.5 mM MgCl₂·6H₂O, 1 mM CaCl₂·2H₂O, 1 ml of vitamin solution and 1 ml of the selenium stock solution in 50 ml nanopure water. The master solution was filter-sterilized and flushed for 1 h and 40 min by 20% CO₂:80% N₂ gas mixture. This solution was added to the FGL medium immediately before inoculation, 5 ml per 100 ml medium. Manganese was added at the time of inoculation from concentrated anaerobic stock solution of MnCl₂·4H₂O (1 M). Finally, after inoculation, the medium was reduced by the addition of sulfide from a concentrated anaerobic stock solution of Na₂S·5H₂O (0.2 M).

The medium used for the initial enrichment was reduced by the addition of 4 mM sodium ascorbate instead of sulfide, to minimize the growth of organisms that require high concentrations of sulfide as an electron donor. A brown microbial mat formed on the surface

of the inoculated sediments after 3–4 weeks. Fragments of this mat were transferred into the sterile medium with the same composition as described above, incubated in the same conditions for one month and transferred again. Most experiments described here used biofilms that had undergone at least four transfers from the initial enrichment.

Modified FGL medium. All experiments described in the main text used the modified basal FGL (MFGL) medium. This medium did not contain DCMU, sulfate or ascorbate, and was reduced by 20–50 µM Na₂S. Sterile MFGL contained only traces of sulfate (<0.9 µM), nitrate (<0.5 µM) and nitrite (<0.1 µM), as detected by ion chromatography (see ‘Concentrations of dissolved species in culture media’).

To evaluate the influence of light on growth, biofilms from the third transfer were inoculated into the MFGL medium. One triplicate set of batch cultures was incubated for 2 weeks in the light at 27 °C at a distance of 35 cm from an incandescent white light bulb that emits between 400–700 nm. Another set was incubated at the same time and at the same temperature but was shielded from the light by aluminium foil.

To reduce the carryover of manganese oxides in the biofilm inoculum, we reduced the inoculums using a previously described protocol³². In brief, microbial biofilms that had been grown in the presence of light and 1 mM Mn(II) were collected and incubated in anaerobic sterile ascorbic acid (0.25 mM) in the anaerobic chamber for 10 min. After this incubation, the biofilms were washed three times with sterile, anaerobic nanopure water. XRD analyses of biofilms treated in this manner showed that this protocol removed manganese oxide minerals but increased the abundance of elemental sulfur in the inoculum. Elemental sulfur was absent from the biofilms before the treatment.

To characterize the effect of different initial concentrations of manganese and sulfide on mineral precipitation, biofilms from the third transfer of the original enrichment culture were inoculated into the MFGL medium amended by MnCl₂ and Na₂S from 0.5 M and 0.1 M anaerobic stock solutions. All stock solutions were prepared, autoclaved and stored under an atmosphere of N₂. The effect of manganese concentration on manganese oxidation was evaluated in three sets of triplicate inoculated cultures that contained 0.1, 1 or 5 mM MnCl₂. All these cultures were reduced with 50 µM Na₂S. The effect of H₂S concentration on manganese oxidation was explored in three sets of triplicate cultures reduced by 0.05, 0.25 or 1 mM of Na₂S, all amended with 1 mM MnCl₂. An additional set of triplicate cultures contained 1 mM Mn(II) and 0.02 mM Na₂S. All cultures were incubated for two weeks.

Further enrichment of manganese-oxidizing and sulfide-oxidizing microorganisms. Microbial communities capable of anaerobic oxidation of manganese were further enriched by inoculating anaerobically sealed agar shake tubes with the dispersed biofilms, serially diluting the cultures in agar, transferring colonies into liquid medium and repeating the entire process for a second time^{33,34}. The MFGL medium in agar shake tubes was solidified by 1.1% agar. Biofilms were washed with anaerobic nanopure water in the anaerobic glove box and mechanically dispersed in 10 ml of the basal MFGL medium. The first agar shake tube was inoculated with 10% (1 ml) of the dispersed inoculum and diluted by 5 successive transfers of 1 ml into 9 ml of sterile MFGL.

The additions of sulfide and manganese to the basal MFGL in agar shake tubes targeted two different conditions: condition (1), 0.02 mM Na₂S and 1 mM MnCl₂, sought to enrich for microorganisms that can photosynthesize in the presence of low sulfide concentrations and oxidize Mn(II); and condition (2), 1 mM Na₂S (MnCl₂ added only in the trace metal solution), enriched for *Chlorobium* spp. that can oxidize sulfide. Extended Data Table 2 summarizes the enrichment protocol and conditions. The shake tubes were incubated at 27 °C at a distance of 35 cm from the incandescent white light bulb. Colonies that formed after one month were transferred from the solid medium into the liquid

medium that contained the same concentrations of MnCl_2 and Na_2S . Biofilms that grew in liquid after one month were mechanically dispersed by a syringe, inoculated into another set of agar shake tubes and incubated for one month. Colonies from the shake tubes were inoculated again into the liquid medium. The purity of the cultures at each transfer was tested by Sanger sequencing. Amplified 16S rRNA genes were sequenced in both directions using either 27F (5'-AGAGTTT-GATCCTGGCTCAG-3') or 1492R (5'-ACG GCT ACC TTG TTA CGA CTT-3') (Integrated DNA Technologies), assembled to get a nearly full-length 16S rRNA gene (GeneWiz) and identified using nucleotide BLAST on GeneBank³⁵. Future experiments should also explore the possibility of light-dependent production of organoperoxides in the medium as a function of vitamins and Fe(II) in the medium.

DNA extraction, 16S rRNA gene Illumina sequencing and phylogenetic analyses. A 500- μl sample of each biofilm enrichment (including early enrichments and the lake inoculum) was collected and spun down into a pellet. Total DNA was extracted from samples using the PowerSoil DNA Isolation Kit (MoBio) according to manufacturer's instructions and eluted in 60 μl C6 solution. Upon extraction, DNA was quantified using NanoDrop (Thermo Scientific). The extracted samples and blank-template controls from the PowerSoil DNA Isolation kit were stored at -80°C and sent to Argonne National Laboratory on dry ice for sequencing. The community composition was characterized using 16S rRNA gene amplicon paired-end sequencing on the MiSeq Illumina platform. In brief, V4 region of the 16S rRNA gene (515F–806R) from each sample was amplified using the bacterial-specific primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806 R (5'-GGACTACHVGGGTWTCTAAT-3')³⁶. After the amplifications, the PCR amplicons were quantified using Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher/Invitrogen, P11496) according to manufacturer's instructions and pooled in equal concentrations (240 ng) to a single tube. This pool was cleaned up using MoBio UltraClean PCR Clean-Up Kit (MoBio) and quantified using the Qubit (Invitrogen). The pooled samples were sequenced on the Illumina MiSeq platform (Illumina). All library preparations, pooling, quality controls and sequencing runs were performed at the Argonne National Laboratory. Sequence data were analysed using QIIME v.1.9.0³⁷. Paired-end reads were joined using fastq-join method³⁸, and libraries were demultiplexed and filtered. Any reads that did not assemble by perfect matches in the overlapping region or meet the q -score (>20) threshold were removed and were not used in subsequent analyses. Chimeric sequences were identified using UCHIME's search61 de novo-based chimera detection algorithm³⁹ and removed from the quality-filtered sequences. Filtered and chimera-free sequences were aligned and clustered into operational taxonomic units at $>97\%$ similarity level using closed-reference UCLUST algorithm against the Greengenes v.13.8 reference dataset as a database⁴⁰. The most abundant sequence from each cluster was selected as a representative sequence. All representative sequences were aligned using PyNAST³⁷. A phylogenetic tree for subsequent phylogenetic analyses was built using FastTree⁴¹. Operational taxonomic unit counts were rarified to 10,000 sequences per sample for diversity analysis using taxonomic and phylogenetic indices that included the Shannon and Faith's PD index. To identify bacterial taxa with sequences that are more abundant in samples grown in light and/or with Mn(II), we used LEfSe, which performs a nonparametric Wilcoxon sum-rank test followed by linear discriminant analysis (LDA), coupled with effect size measurements to assess differentially abundant taxa⁴². *Chlorobium* sp. sequences were significantly enriched in samples grown in the presence of light and Mn(II) with LDA scores >5 . Cultures grown in light had significantly more *Chlorobium* sp. (analysis of variance (ANOVA), $F = 23.4521$, df factor = 5, df error = 12, $P < 0.0001$; Tukey's HSD, $P < 0.01$). Sequence data are available as FASTQ files at the National Center for Biotechnology Information (NCBI) via Sequence Read Archive (SRA), under the SRA accession ID number SRP133329.

Metagenome sequencing and analysis. To determine the metabolic potential of cultures grown from colonies that targeted specific growth conditions (see 'Further enrichment of manganese-oxidizing and sulfide-oxidizing microorganisms'), we sequenced their metagenomes. The DNA of enrichments obtained using condition (1) was extracted using a modified phenol–chloroform method with ethanol precipitation as previously described⁴³ and quantified by a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). This DNA was sent for metagenomic sequencing to the University of Southern California's Genome and Cytometry Core Facility. The library preparation, quality control and sequencing were performed at the Cytometry Core Facility. In brief, before sequencing on the Illumina HiSeq 2500 platform, DNA was sheared using dsDNA ShearAs Plus (Zymo), cleaned up using Agencourt AMPure XP beads (Beckman-Coulter), the library was quantified using the Qubit 2.0 Fluorometer and the DNA fragment size was determined with an Agilent Bioanalyzer 2100.

The quality control of the sequence data was performed using Trimmomatic v.0.36 using default parameters and a minimum sequence length⁴⁴ of 36 bp. IDBA-UD v.1.1.2. was used to assemble the reads with a 2,000 bp minimum contig length. SAMtools v.1.3.1⁴⁵ was used to convert files to binary format for downstream analysis. VizBin was used to delineate individual genomes from the enrichment metadata⁴⁶ and the genomes were assigned putative taxonomic identities according to their placement in a phylogenetic tree in CheckM v.1.0.4 using the 'tree' command⁴⁷.

Individual genomes obtained from the metagenome data were submitted to the DOE Joint Genome IMG-MER (Integrated Microbial Genomes) pipeline for gene calling and assembly⁴⁸. The protein-coding-gene prediction tool Prodigal v.3.0.0 was used to determine genes in the enrichment grown from the colony on 1 mM MnCl_2 and 20–50 μM Na_2S . The genome of *C. limicola* was 98.4% similar to *C. limicola* Frassasi⁴⁹.

To detect putative Mn(II)-oxidizing genes in *C. limicola*, we first generated a blast database of protein-coding Mn(II)-oxidizing genes by selecting genes that encode for multi-copper oxidases and animal haem peroxidases. Because multi-copper oxidases and animal haem peroxidases each contain several classes of enzymes and can transfer electrons from a number of different substrates, we focused on enzymes with confirmed manganese-oxidizing activities by biochemical and molecular assays. All multi-copper oxidases and animal haem peroxidases involved in Mn(II) oxidation and characterized to date are from aerobic microorganisms and include genes such as *mnxG*, *mcoA* and *mopA* in *Pseudomonas putida*¹⁸, *mnxG* in the spores of *Bacillus* strain SG-1¹⁹, *moxA* in *Pedomicrobium* sp. ACM 3067⁵⁰, and *mopA* in *Aurantimonas manganoxydans* S185-9A1⁵¹ and *Roseobacter* sp. AzwK-3b¹⁹. To determine whether *Chlorobium* has any homologues with characterized manganese-oxidizing multi-copper oxidases and animal haem peroxidases, we used BLASTp⁵² and queried translated Mn(II)-oxidizing genes against the *Chlorobium* genome with an e -value cut-off of 10^{-5} and a bit score of 30. Homologues of multi-copper oxidases and animal haem peroxidases in *C. limicola* are shown in Extended Data Table 4.

Sequence data for *C. limicola* can be accessed at the JGI-IMG under IMG submission ID 124328.

Spectroscopy

X-ray photoelectron spectroscopy. X-ray photoelectron spectroscopy (XPS) was performed on a K-alpha + X-ray photoelectron spectrometer (K-Alpha + XPS, Thermo Fisher). Biofilms were collected and centrifuged at 14,000 rpm for 5 min in the anaerobic chamber to form pellets. The pellets were placed on double-sided carbon tape and dried in the anaerobic chamber. To maintain the anoxic conditions, the samples were stored in the anaerobic chamber in hermetically sealed glass vials before analysis. All samples were fractured in high vacuum (3×10^{-8} Torr) in the outer pressure chamber and then moved directly

into the main XPS measurement chamber. An incident monochromatic X-ray beam from the Al K Alpha target (15 kV, 10 mA) was focused on a 0.4 mm × 0.3-mm area of the surface at a 45° angle with respect to the sample surface. Depth profile etching with an etch cycle of 30 s and a total of 10 levels yielded high-resolution spectra. The electron energy analyser perpendicular to the sample surface was operated with a pass energy of 50 eV to obtain XPS spectra at a 0.1-eV step size and a dwell time of 50 ms. Each peak was scanned 15 times. To ensure representative data from heterogeneous samples, we probed a total of 50–80 points per sample. XPS data were treated and analysed using CasaXPS curve resolution software package. Spectra were best fit after Shirley background subtractions by nonlinear least squares CasaXPS curve resolution software package. Gaussian and Lorentzian contributions to the line shapes were numerically convoluted using a Voigt function. The different XPS lines with sets of Gaussian and Lorentzian peaks were empirically fitted with different standards corresponding to different oxidation sets (MnO, MnCO₃, Mn₂O₃, Mn₃O₄, MnO₂ and MnCaO₃). Each manganese XPS spectrum was empirically best fitted with multiple standard phases (MnO, MnCO₃, Mn₂O₃, Mn₃O₄, MnO₂ and MnCaO₃) that produced the minimum residual. The average fit properties for all treated spectra were acceptable as the following: $R_{\text{expected}} = 1.60$, $R_{\text{profile}} = 1.71$, significance level = 0.05, residual s.d. = 1.67, goodness of fit = 1.78, critical $\chi^2 = 3.84$.

Interpretation of XPS spectra. The redox state of manganese in microbial cultures was confirmed by XPS (Extended Data Fig. 3). The Mn2p XPS spectra of the dark culture exhibited two major peaks at binding energies of 640.90 eV and 652.2 eV, which correspond to Mn2p_{2/3} and Mn2p_{1/2}, respectively. This is consistent with other reports^{53,54} on Mn(II) phases of Mn. In the photosynthesizing culture, the Mn2p peak shifted to a high-energy side and the intense satellite peak characteristic of Mn(II) diminished. These biofilms contained manganese in different valence states. At some analysed spots, the Mn2p XPS spectrum exhibited two major peaks of Mn2p_{2/3} and Mn2p_{1/2} at binding energies of 642 eV and 653 eV, respectively. These correspond to Mn(IV) in calcium–manganese oxide phases⁵⁵. Peaks at Mn2p_{2/3} with binding energies 641.61 eV and 641.47 eV, respectively, were also detected. These peaks correspond to Mn(III) in Mn₂O₃ and Mn(III) and Mn(II) in Mn₃O₄ phases^{56,57}.

The redox state of the manganese in the culture enriched in condition (1) and condition (2) was confirmed by XPS (Extended Data Fig. 4). The Mn2p XPS spectra of this culture (Extended Data Fig. 4a) exhibited two major peaks at binding energies of 641.41 eV and 653.15 eV, corresponding to Mn2p_{2/3} and Mn2p_{1/2}, respectively, and matching the peaks^{58,59} of Mn₃O₄. The Mn2p XPS spectra of the condition (2) enrichment (Extended Data Fig. 4b) exhibited Mn2p_{2/3} peaks at 640.97 eV and 652.2 eV, which correspond to Mn2p_{2/3} and Mn2p_{1/2}, respectively. This is consistent with other reports of MnO phases^{56,58}.

Probing the redox state of manganese in biofilms. We used XPS to detect oxidized manganese in colonies enriched on 1 mM Mn(II) (condition (1)). Extended Data Table 3 summarizes the procedure used to study the Mn(II) oxidation activity in the enrichment cultures. Manganese oxidation was tested using cultures that were enriched as colonies in agar shake tubes (condition (1) and condition (2)) (see Extended Data Table 3 and ‘Spectroscopy’ and ‘Interpretation of XPS spectra’). Biofilms from condition (1) were grown in duplicate 10-ml cultures with 1 mM MnCl₂ and 0.02 mM Na₂S, mechanically dispersed and resuspended into separate 10-ml liquid solutions. Five per cent v/v of this suspension was transferred into 10 ml of MFGl medium with 1 mM MnCl₂ and 0.02 mM Na₂S and the cultures were incubated at 27 °C for 1 week before the assay. A second assay tested the Mn(II) oxidation activity without requiring the very sparse biofilm to grow. The condition (1) enrichment was grown for two weeks as described in ‘Probing the redox state of manganese in biofilms’, centrifuged at 8,000 rpm in the anaerobic

chamber, washed 3 times with anaerobic water and transferred into 10 ml of MFGl with 1 mM MnCl₂ and 0.02 mM Na₂S. These cultures were incubated for 3 days in a 12:12 h light:dark regime and collected anaerobically. This procedure preserved the cell density of the original biofilms and did not require microbial growth.

To test for Mn(II) oxidizing activity in the enrichment from condition 2 (*Chlorobium* sp. and *Desulfomicrobium* sp.), the culture was grown in 10 ml of the basal MFGl amended with 1 mM Na₂S, collected anaerobically and dispersed in 10 ml of the basal MFGl medium. This suspension was used to inoculate 10 ml of the basal MFGl amended with 1 mM MnCl₂ and 0.02 mM Na₂S at 5% v/v. The inoculated medium was incubated in the light/dark regime for one week. To test for Mn(II) oxidizing activity without requiring the low-biomass biofilms to grow, the enrichment from condition (2) was grown for 2 weeks in 10 ml of the basal MFGl amended with 1 mM Na₂S, centrifuged at 8,000 rpm in the anaerobic chamber, washed 3 times with anaerobic water and transferred to 10 ml of MFGl with 1 mM MnCl₂ and 0.02 mM Na₂S. These cultures were incubated for 3 days in the 12:12 h light:dark regime and collected anaerobically. All collected microbial pellets were dried on carbon tape and stored anaerobically inside serum bottles with N₂ atmosphere and placed inside the anaerobic chamber in the dark at 26 °C until XPS analysis.

Probing the redox state of manganese in co-cultures. *C. limicola* (DSM 245, DSMZ) and *C. tepidum* (DSM 12025, DSMZ) were inoculated with 5% v/v inoculum and grown in 50 ml MFGl medium supplemented with 0.05 mM Na₂S and 0.5 g/l yeast extract in a 12:12 h light:dark regime at 27 °C for 3 weeks. *G. lovleyi* (DSM 17278, DSMZ) was grown in MFGl supplemented with 2.8 mM ferrihydrite and 5 mM acetate and reduced with 0.05 mM Na₂S in the dark at 21 °C for 3 weeks. Microorganisms from these cultures were inoculated as 5% v/v inoculum in the following combinations: *C. limicola* + *C. tepidum*, *C. tepidum* + *G. lovleyi*, *C. limicola* + *G. lovleyi* and *C. limicola* + *C. tepidum* + *G. lovleyi*. All these co-cultures were grown in 10 ml MFGl medium with 1 mM Mn(II) and 0.05 mM Na₂S in a 12:12 h light:dark regime at 27 °C for 2 weeks. The biomass was collected anaerobically, the pellets were dried on carbon tape and stored anaerobically under N₂ inside the anaerobic chamber in the dark at 26 °C until XPS analysis. The oxidation state of manganese was characterized by XPS in all four cultures.

X-ray powder diffraction

XRD patterns were obtained in reflection mode with Ni-filtered Cu K α radiation ($\lambda = 1.5406 \text{ \AA}$) as X-ray source on an X’Pert PRO diffractometer (XRD, X’Pert PRO, PANalytical) equipped with an X’Celerator detector (PANalytical). The patterns were measured in 2θ range from 3° to 90° with a scanning step of 0.008° and a fixed counting time of 600 s at 45 kV and 40 mA. Biofilms were collected and centrifuged at 14,000 rpm for 5 min in the anaerobic chamber. Microbial paste was smeared on Zero Diffraction Disk (23.6-mm diameter × 2.0-mm thickness, Si crystal, MTI) and dried in the anaerobic chamber. The samples were analysed inside the anaerobic dome to maintain the anoxic conditions during the XRD analyses. Data were analysed and fitted using High Score Plus program version 4.5. The average fit properties for all treated spectra were acceptable as the following: residual s.d. = 1.63, $R_{\text{expected}} = 1.28$, $R_{\text{profile}} = 1.63$, significance level = 0.05, goodness of fit = 1.69, critical $\chi^2 = 3.84$.

Precipitated minerals were also analysed using in-situ synchrotron-based XRD at the Advanced Light Source at the beamline 12.3.2. Biofilms were collected on site and the biofilm paste was loaded into transmission sample XRD cells. The transmission synchrotron diffraction data were collected using a DECTRIS Pilatus 1M hybrid pixel area detector placed at an angle 2θ of 35° at approximately 170 mm from the sample. The 4-bounce monochromator was set to an energy was 10 keV ($\lambda = 1.239842 \text{ \AA}$). The sample geometry with respect to the incident beam and the detector was carefully calibrated using Al₂O₃ powder. The 2D diffraction patterns (Fig. 1d) were analysed and integrated along the

Article

azimuthal direction into 1D diffractograms using the X-ray microdiffraction analysis software (XMAS v6) developed at the Advanced Light Source for the Advanced Light Source beamline 12.3.2, and Matlab R2017a.

Determination of XRD detection limit. To determine the detection limit of XRD, 0.05, 0.01, 0.02, and 1 mg of MnO_2 was mixed with 10 mg of dry anaerobic biofilm that did not contain green sulfur bacteria or manganese oxides and spread on Zero Diffraction Disk (23.6-mm diameter \times 2.0-mm thickness, Si crystal, MTI). The mixtures were analysed by X'Pert PRO diffractometer XRD, X'Pert PRO, PANalytical) equipped with an X'Celerator detector (PANalytical) over 10-h analysis time. MnO_2 standard and the bacterial biofilm were also run separately as controls. The detection limit of XRD was determined by the mass of MnO_2 that yielded discernible diffraction peaks in the XRD spectrum.

Interpretation of XRD peaks. The XRD spectra of microbial cultures incubated in the light with Mn(II) (Fig. 2) showed peaks that can be indexed to a ternary manganese oxide; CaMnO_3 (ICDD-01-016-2217) with lattice constants of $a = 5.2917$ nm, $b = 7.4803$ nm and $c = 5.2870$ nm⁵⁸. CaMnO_3 is not known to occur naturally.

Dolomite was the most abundant phase in the cultures and its peaks were indexed as (104), (101), (110), (11-3), (202) and (018) (ICDD-04-011-9833). The absence of light inhibited the growth of photosynthetic microorganisms and the formation of manganese-oxide minerals, and also reduced the precipitation of dolomite (Fig. 2). Biofilm incubated in the dark showed the precipitation of calcium carbonate phase, CaCO_3 (ICDD-00-058-0471) indexed for (121) and (102). In addition to the various carbonate phases, the XRD spectrum showed two different phases of elemental sulfur (S^0); (ICDD-04-020-2294) indexed for (110), (-101) and (-211) and (ICDD-05-001-0219) indexed for (110) and (-101).

The XRD spectra of microbial cultures incubated at different concentrations of manganese and sulfur showed peaks of manganese oxide, dolomite and elemental sulfur (Extended Data Fig. 2). The latter formed in microbial cultures that were incubated with 0.25–1 mM H_2S and in the cultures grown with less than 1 μM Mn(II) (see the composition of the trace metal solution in 'Enrichment and culturing conditions'). S^0 (ICDD-05-001-0219), was indexed for (110), (-101), (011) and (-211).

Microscopy

Scanning electron microscopy. Scanning electron micrographs were acquired by a Zeiss Merlin scanning electron microscope with the GEMINI II column (Zeiss Merlin SEM, Carl Zeiss microscopy). The microscope was equipped with a field gun emission and energy-dispersive X-ray spectrometer (EDS, EDAX detector; EDAX) that operated at an accelerating voltage of 5–15 kV, probe current of 100 pA and a working distance of 8.5 mm. On-axis in-lens secondary electron detector was used during imaging. The samples were fixed by 0.2 M sodium cacodylate, 0.1% CaCl_2 and 2.5% glutaraldehyde in anaerobic water for 2–3 days at 4 °C. The fixed samples were washed by 0.1 M sodium cacodylate followed by a wash in nanopure water. After washing, the samples were dehydrated in a series of ethanol–water solutions. The ethanol–water solution series included the following dehydration steps: 30% (20 min), 50% (20 min), 70% (20 min), 80% (20 min), 90% (20 min) and 100% (3 \times 20 min) of 200 proof ethanol. After air-drying, the samples were mounted on double-sided carbon tape and coated with a thin layer 5 nm of Au/Pd or 10 nm of carbon using a Hummer V sputter coater. EDS spectra were treated and analysed by TEAM EDS 2.0 analysis software (EDAX) and Microsoft Excel 2016.

TEM. Transmission electron micrographs were obtained using FEI Tecnai F20 supertwin microscope (FEI Tecnai G2, FEI) with a 200-kV Schottky field emission gun. The samples were imaged at 80 kV with 1,024 \times 1,024 CCD Gatan camera (Gatan). The samples were fixed by 0.2 M sodium cacodylate, 0.1% $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$ and 2.5% glutaraldehyde

in aerobic nanopure water for 2–3 days at 4 °C. The samples were then washed with washing buffer (0.1 M sodium cacodylate in nanopure water), postfixed with 1% osmium tetroxide in water for 1 h, washed with aerobic nanopure water and stained with 1% uranyl acetate for 1 h. The stained samples were washed with nanopure water and dehydrated in an ethanol–water solution series. The ethanol–water solution series included the following dehydration steps: 30% (20 min), 50% (20 min), 70% (20 min), 80% (20 min), 90% (20 min), and 100% (3 \times 20 min) of 200 proof ethanol. The samples were further dehydrated with propylene oxide:ethanol solvent (50:50, by vol) for 30 min, then with 100% propylene oxide. The epoxy resin used for embedding consisted of diglycerol ether of polypropylene glycol (EmBed 812, DER 736, Electron Microscopy Sciences, EMS no. 14130), cycloaliphatic epoxide resin (ERL 4221 Electron Microscopy Sciences, EMS no. 14300), nonenyl succinic anhydride (NSA, Electron Microscopy Sciences, EMS no. 14300) and 2-(dimethylamino)ethanol (DMAE, Electron Microscopy Sciences, EMS no. 14300). The samples were embedded in resin and cut into 80-nm-thick sections with a diamond knife using Leica Reichert Ultracut E microtome (Reichert Ultracut E microtome, Leica) with a thickness setting of 50 nm. Thin sections were placed on FCF-200 grids (Electron Microscopy Sciences, FCF-200-Cu).

To determine whether the fixation and embedding protocols introduced any artefacts, photosynthetic biofilms were also collected without any further processing or staining in the anaerobic chamber. A drop of microbial culture was deposited on LC-200 grid (Electron Microscopy Sciences, LC-200-Cu) and imaged with JEOL 2010F TEM (JOEL 2010F, JOEL). The JEOL 2010F TEM is equipped with a Schottky field emission gun operating at 200 kV and a Gatan energy filter (GIF, Gatan 200, Gatan). The 2010F TEM has micro-diffraction, diffraction pattern in parallel beam and convergent beam electron diffraction features to allow SAED on selected mineral-encrusted bacteria with a high spatial resolution. Gold standard was used as reference for SAED analyses. The high-angle annular dark field detector (Gatan) for atomic resolution scanning electron transmission microscopy in the free-lens control mode (STEM) and with an EDS (Bruker silicon drift detector SDD, Bruker) enabled elemental analysis at nanoscale resolution. Images in the TEM and STEM mode were taken by a digital camera (Gatan Orius, Gatan). SAED patterns were imaged using Gatan digiscan unit (Gatan). TEM, STEM and SAED images were recorded and treated using Gatan digital micrograph software (Gatan). EDS spectra were recorded and treated using INCA program (Oxford Instruments).

Interpretation of SAED patterns. Different types of manganese minerals in photosynthetic biofilms corresponded to different stages of mineral maturation. High-resolution TEM of the manganese oxide nanocluster surrounding a cell (Fig. 3) showed polycrystalline minerals with a uniform lattice fringe that corresponded to the (116) plane with interplanar spacing of 2.71 Å of calcium manganese oxide (ICDD-00-053-0092). The SAED patterns of minerals that were not associated with cell surfaces showed various minerals. One type of manganese mineral had four obvious polycrystalline diffraction rings that could be observed at 3.65 Å, 3.40 Å, 2.88 Å and 1.83 Å. These corresponded, respectively, to the (112), (211), (220) and the (323) crystal planes of Mn_3O_4 (ICDD-03-065-2776) (Fig. 3). Some globular nanocrystals of manganese oxide outside of any microbial surfaces (Fig. 3) showed lattice fringes with the interplanar spacing of 2.26 Å. This matched the characteristic interplanar spacing of the (200) plane of manganese oxide type MnO mineral (ICDD-04-004-3858).

Concentrations of dissolved species in culture media

Sulfide concentrations were determined using a modification of a previously published method⁶⁰ in samples of triplicate cultures for each time point. In brief, 200 μl of each liquid sample was diluted in 1 ml of 0.05 M zinc acetate. Standards were prepared from 1 mM anaerobic stock solution of Na_2S diluted by 0.05 M zinc acetate. The concentration of Na_2S

stock solution was verified by precipitating an exact volume of Na₂S with an excess volume of 0.3 M silver nitrate. Six hundred microlitres of the precipitated sample were transferred and reacted with 10 µl of diamine reagent. After 20 min reaction time in the dark, the absorbance was measured by a multi-mode reader spectrophotometer (BioTek, Synergy 2) at 670 nm.

The concentrations of sulfate, nitrite and nitrate in the samples of the liquid medium from triplicate cultures were determined by ion chromatography (Dionex ICS-16000 equipped with an auto-sampler Dionex AS-DV, Thermo Fisher), guard column (Dionex Ion Pac AG22, RFIC, Guard 2 × 50 mm, Thermo Fisher), analytical column (Dionex Ion Pac AS22, RFIC, Analytical 2 × 250 mm, Thermo Fisher, USA) and a trap column for metals (Dionex Ion Pac MFC-1, RFIC, trap column, metal free, 3 × 27 mm, Thermo Fisher). All samples were filtered anaerobically through 0.2-µm-pore-size filters (Acrodisc 25-mm syringe filter, PALL) and stored at −20 °C. The chloride ion was solid-phase extracted from all samples using a Ag/H cartridge (Dionex OnGuard II Ag/H, 2.5 c.c. cartridge, Thermo Fisher) before the analysis. The removal of the chloride ion affected the lower detection limit for phosphate, but not for sulfate and nitrate. The limits of detection for sulfate, nitrate and nitrite, respectively, were 20 µg/l, 20 µg/l and 10 µg/l respectively.

Total dissolved manganese concentrations in the liquid culture medium from triplicate cultures were determined by inductively coupled plasma-mass spectrometry (ICP-MS, Agilent 7500, Agilent). All samples were filtered through 0.2-µm-pore-size filters (Acrodisc 25-mm syringe filter, PALL), and acidified with 2% high purity HCl (hydrochloric acid 30%, Sigma Aldrich, suprapur- end Millipore, 100318) and stored at −20 °C. All samples were diluted with high-purity 2% HCl (hydrochloric acid 30%, Sigma Aldrich, suprapur- EMD Millipore, 100318) before the analysis.

Dissolved manganese in the liquid phase was also measured by leucobertelin blue (LBB) assay⁶¹ and iodometric method⁶². Oxidized manganese in the liquid phase includes any soluble valence state of manganese that can pass through the 0.2-µm pore filter. We used the iodometric method^{62,63} to determine manganese oxidation state including Mn(II), Mn(III), Mn(IV) and Mn(VII). Again, none of the measurements detected oxidized manganese in the liquid phase. The LBB assay was also used to quantify the concentrations of oxidized manganese in biofilms. These concentrations were detected as oxidizing equivalents of KMnO₄ ('Quantification of biofilms and oxidized manganese in biofilms').

The concentration of peroxide was measured in triplicate samples of microbial biofilms and sterile controls incubated in the light with 1 mM Mn(II) and 50 µM Na₂S using a peroxidase activity assay kit (Sigma Aldrich, MAK092). The standard curve was measured using different dilutions of the H₂O₂ standard (Sigma Aldrich, MAK092C) in sterile culture medium mixed with the reaction mix composed of 2 µl fluorescent peroxidase substrate (Sigma Aldrich, MAK092B) and 48 µl of HRP positive control (Sigma Aldrich, MAK092D). A 100 µl of each diluted H₂O₂ standard and the samples were distributed into microplate wells. The plate was incubated at 37 °C and the initial measurement (t_{initial}) was measured after 3 min by multi-mode reader spectrophotometer (BioTek, synergy 2) at 570 nm. The absorbance was measured every 3 min until the value of the most-active sample exceeded the end linear range of the standard curve. We did not detect any H₂O₂ in the incubations or sterile controls. The limit of detection of H₂O₂ using colorimetric detection was 0.1 nM.

Quantification of biofilms and oxidized manganese in biofilms

The amount of biofilm was measured by crystal violet staining using a modification of a previously published assay⁶⁴. In brief, biofilms from triplicate serum bottles were collected at each time point and centrifuged aerobically at 10,000 rpm for 30 min. The supernatant was decanted and 0.5 ml of 0.1% of aqueous crystal violet added (crystal violet, Sigma Aldrich, ACS reagent, ≥ 90% anhydrous basis, C6158). The stained biofilm was incubated in the dark at room temperature for 24 h,

washed 15 times with nanopure water and air-dried. After drying, 0.5 ml of 30% acetic acid (37% acetic acid, Sigma Aldrich, ACS reagent, ≥ 99.7%, 695092) was added to the samples and left to react at room temperature for 30 min. Acetic acid solubilized all crystal violet molecules bound to peptidoglycan and exopolysaccharide. Thus, the solubilized crystal violet corresponds to the biomass in biofilms. The collected solubilized crystal violet was filtered through 0.2-µm-pore-size filters (Acrodisc 25-mm syringe filter, PALL) and 200 µl of the solution was transferred into a microtiter plate. The absorbance of the samples was measured at 550 nm using a spectrophotometer (BioTek, synergy 2).

Oxidized manganese in biofilms was quantified by the LBB assay. Biofilms were inoculated from frozen stocks into triplicate serum bottles that contained 25 ml or 50 ml of MFL medium and incubated for one, two or three weeks. Biofilms from the frozen stock (about 5 mg) were washed 10 times with anaerobic nanopure water to remove any glycerol, inoculated into the culture medium and the medium was immediately flushed with 20% CO₂/80% N₂ (v/v) for 1 h. The biofilms grew for 2 weeks in the light, at which point, biofilms (about 5 mg) were transferred into serum bottles that each contained 25 ml of the fresh MFL medium. Three bottles were incubated in the light, three in the dark and all biofilms were collected after one or two weeks by pipetting and centrifugation in the anaerobic glove box. After the LBB assay, all analysed samples were air-dried for >24 h and weighed. The one-week old biofilms weighed 16–18 mg; the two-week old biofilms weighed 19–21 mg. A separate experiment quantified the amount of oxidized manganese in duplicate 25-ml cultures of 3-week old biofilms that had been inoculated with about 0.1 mg of the washed material from frozen culture stocks and weighed <0.3 mg at the end of the experiment. In contrast to the experiments that yielded samples for XRD, XPS, scanning electron microscopy and TEM analyses, these experiments involved at most one successive transfer of biofilms after the inoculation from frozen stocks of biofilms enriched as described in 'Enrichment and culturing conditions'.

The working reagent was prepared as 0.04% LBB in 45 mM acetic acid and stored at 4 °C overnight in a light-proof container. Potassium permanganate (KMnO₄) 1 mM stock solution was freshly prepared in water and standards (5, 10, 15, 20, 40 and 50 µM) were prepared by diluting the stock solution in water. The samples were incubated for 20 min in 0.75 ml of the 0.04% LBB working reagent in the dark at room temperature and centrifuged for 90 s at 10,000 rcf to remove the biofilm and mineral particulates from the solution. The absorbance of the supernatant was measured on a spectrophotometer at 618 nm. To determine whether some manganese was oxidized in the dark, about 5 mg of the biofilm stock was inoculated into sterile MFL in the dark for 2 weeks which detected on average 0.02 µM oxidizing equivalent per 5 mg of biofilm. Control experiments assayed the concentration of oxidized manganese in FGL enrichment cultures that contained 1 mM sulfide, and 1 mM MnCl₂ and did not detect any oxidized manganese.

Oxygen concentration

To determine how much oxygen can diffuse into the cultures through the butyl rubber caps, we used our in-house-developed oxygen sensor⁶⁵ based on the fluorescence lifetime⁶⁶ of 5,10,15,20-tetrakis(pentafluorophenyl)-21H,23H-porphinepalladium (II). The sensor can detect changes in the partial pressure of oxygen that are smaller than 1 µbar and its main sensitivity region⁶⁵ is 0–100 µatm.

Experiments were conducted to quantify the oxygen concentration in the cultures and the maximum amount of oxygen inflow. First, the partial pressures of oxygen in the headspaces of photosynthetic cultures, sterile controls incubated in the light and dark control cultures were measured automatically for 14 days. All serum bottles contained 100 ml of the medium reduced by 50 µM Na₂S (Extended Data Fig. 1). The partial pressure of oxygen in the headspaces of the bottles did not increase or fluctuate by more than 2 µbar over the course of the growth experiment. The main sources of noise were daily thermal fluctuations

(high-frequency component) and sensor aging (low-frequency component). The upper limit for the partial pressure of oxygen in the headspace is 2 μ bar, measured in the beginning of the experiment. This partial pressure was lower than 0.5 μ bar during most of the experiment. This corresponds to a maximum dissolved molecular oxygen concentration of 2.6 nM, assuming the equilibrium between O₂ in the headspace and O₂ dissolved in the culture medium according to Henry's law.

In an additional test, we incubated biofilms in the light in the anaerobic chamber under a 5%CO₂/5%H₂/balN₂ (v/v/v) atmosphere. The partial pressure of oxygen in the chamber was below 1 ppm, as opposed to 21% above the butyl rubber stoppers of the cultures that were incubated outside of the chamber. Therefore, orders-of-magnitude-less oxygen is expected to diffuse into the cultures. The biofilms were grown with and without the addition of 1 mM Mn(II) and the culture medium was reduced with 20 μ M Na₂S. After two weeks of incubation, the biofilms were collected and analysed by XRD. Manganese oxides and carbonates phases formed in biofilms incubated with Mn(II), and elemental sulfur formed in biofilms grown without Mn(II). The formation of detectable quantities of manganese oxides in photosynthetic cultures incubated in the anaerobic glove box further demonstrated the negligible role of oxygen diffusion in the oxidation of manganese.

Acquisition of phototrophy in green sulfur bacteria

Phototrophy within stem-group green sulfur bacteria (GSB) and stem-group green non-sulfur bacteria (GNS) could have been acquired at any point before their post-GOE diversification events. Without additional information, it is not possible to infer where along these branches phototrophy was acquired, but the evolutionary history of bacteriochlorophyll biosynthesis may provide a strong clue. Phylogenies of protein families involved in bacteriochlorophyll biosynthesis have a complex evolutionary history across phototrophic lineages, including gene duplications within the stem-group GSB, and multiple horizontal gene transfer events between GSB and GNS lineages⁶⁷. Specifically, the genes encoding BchH and BchM were transferred from within crown-group GNS to stem-group GSB, with the gene encoding BchH undergoing a duplication shortly before crown-group GSB. BchI is also observed to duplicate in the GSB stem, with one paralogue being transferred to stem GNS. These observations indicate that phototrophy must have existed in these lineages at the time of any bacteriochlorophyll-synthesis-gene duplications, or any divergence of a horizontal-gene-transfer donor lineage. A substantial history of phototrophy within the GSB stem lineage can be inferred from these events (Extended Data Fig. 6). Future molecular clock studies that include these gene tree histories may be able to constrain the time interval for phototrophy in the GSB stem; but the Bch protein histories alone suggest that phototrophy within GSB existed earlier than the appearance of the GSB or GNS crown groups.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence data are available as FASTQ files at the National Center for Biotechnology Information (NCBI) via Sequence Read Archive (SRA), under the SRA accession number SRP133329. The datasets that support the findings of this study are available from the figshare repository (<https://figshare.com/>), with the identifiers 10.6084/m9.figshare.9738515, 10.6084/m9.figshare.9738725, 10.6084/m9.figshare.9738776, 10.6084/m9.figshare.9738905, 10.6084/m9.figshare.9738797, 10.6084/m9.figshare.9738878, 10.6084/m9.figshare.9738887, and 10.6084/m9.figshare.9738896. All other supporting data that support the findings of this study are available from the corresponding authors.

29. Balch, W. E. & Wolfe, R. S. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. *Appl. Environ. Microbiol.* **32**, 781–791 (1976).
30. Sim, M. S., Bosak, T. & Ono, S. Large sulfur isotope fractionation does not require disproportionation. *Science* **333**, 74–77 (2011).
31. Sim, M. S., Ono, S., Donovan, K., Templer, S. P. & Bosak, T. Effect of electron donors on the fractionation of sulfur isotopes by a marine *Desulfovibrio* sp. *Geochim. Cosmochim. Acta* **75**, 4244–4259 (2011).
32. Liang, J., Bai, Y., Men, Y. & Qu, J. Microbe–microbe interactions trigger Mn(II)-oxidizing gene expression. *ISME J.* **11**, 67–77 (2017).
33. Ljungdahl, L. & Wiegand, J. in *Manual of Industrial Microbiology and Biotechnology* (eds Demain, A. L. & Solomon, N. E.) 115–127 (American Society for Microbiology, 1986).
34. Pfennig, N. *Rhodocyclus purpureus* gen. nov. and sp. nov., a ring-shaped, vitamin B12-requiring member of the family Rhodospirillaceae. *Int. J. Syst. Evol. Microbiol.* **28**, 283–288 (1978).
35. Benson, D. A. et al. GenBank. *Nucleic Acids Res.* **41**, D36–D42 (2013).
36. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).
37. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
38. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J.* **7**, 1–8 (2013).
39. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
40. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
41. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
42. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
43. Momper, L. M., Reese, B. K., Carvalho, G., Lee, P. & Webb, E. A. A novel cohabitation between two diazotrophic cyanobacteria in the oligotrophic ocean. *ISME J.* **9**, 882–893 (2015).
44. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
45. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
46. Laczny, C. C. et al. VizBin – an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3**, 1 (2015).
47. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
48. Huntemann, M. et al. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand. Genomic Sci.* **10**, 86 (2015).
49. Mansor, M. & Macalady, J. L. Draft genome sequence of lampenflora *Chlorobium limicola* strain Frasassi in a sulfidic cave system. *Genome Announc.* **4**, e00357-16 (2016).
50. Ridge, J. P. et al. A multicopper oxidase is essential for manganese oxidation and laccase-like activity in *Pedomicrobium* sp. ACM 3067. *Environ. Microbiol.* **9**, 944–953 (2007).
51. Anderson, C. R. et al. Mn(II) oxidation is catalyzed by heme peroxidases in “*Aurantimonas manganoxydans*” strain SI85-9A1 and *Erythrobacter* sp. strain SD-21. *Appl. Environ. Microbiol.* **75**, 4130–4138 (2009).
52. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
53. Nesbitt, H. & Banerjee, D. Interpretation of XPS Mn(2p) spectra of Mn oxyhydroxides and constraints on the mechanism of MnO₂ precipitation. *Am. Mineral.* **83**, 305–315 (1998).
54. Oku, M., Hirokawa, K. & Ikeda, S. X-ray photoelectron spectroscopy of manganese–oxygen systems. *J. Electron Spectrosc. Relat. Phenom.* **7**, 465–473 (1975).
55. Han, X., Zhang, J., Du, F., Cheng, J. & Chen, J. Porous calcium–manganese oxide microspheres for electrocatalytic oxygen reduction with high activity. *Chem. Sci. (Camb.)* **4**, 368–376 (2013).
56. Audi, A. A. & Sherwood, P. M. A. Valence-band X-ray photoelectron spectroscopic studies of manganese and its oxides interpreted by cluster and band structure calculations. *Surf. Interface Anal.* **33**, 274–282 (2002).
57. Toupin, M., Brousse, T. & Bélanger, D. Charge storage mechanism of MnO₂ electrode used in aqueous electrochemical capacitor. *Chem. Mater.* **16**, 3184–3190 (2004).
58. Foord, J., Jackman, R. & Allen, G. An X-ray photoelectron spectroscopic investigation of the oxidation of manganese. *Philos. Mag. A Phys. Condens. Matter Defects Mech. Prop.* **49**, 657–663 (1984).
59. Božin, E. S. et al. Structure of CaMnO₃ in the range 10K ≤ T ≤ 550 K from neutron time-of-flight total scattering. *J. Phys. Chem. Solids* **69**, 2146–2150 (2008).
60. Cline, J. D. Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnol. Oceanogr.* **14**, 454–458 (1969).
61. Krumbein, W. E. & Altman, H. J. A new method for the detection and enumeration of manganese-oxidizing and -reducing microorganisms. *Helgol. Wiss. Meeresunters.* **25**, 347–356 (1973).
62. Murray, J. W., Balistrieri, L. S. & Paul, B. The oxidation state of manganese in marine sediments and ferromanganese nodules. *Geochim. Cosmochim. Acta* **48**, 1237–1247 (1984).
63. Anschutz, P., Dedieu, K., Desmazes, K. & Chaillou, G. Speciation, oxidation state, and reactivity of particulate manganese in marine sediments. *Chem. Geol.* **218**, 265–279 (2005).
64. O'Toole, G. A. Microtiter dish biofilm formation assay. *J. Vis. Exp.* **47**, e2437 (2011).
65. Pajusalu, M., Borlina, C. S., Seager, S., Ono, S. & Bosak, T. Open-source sensor for measuring oxygen partial pressures below 100 microbars. *PLoS ONE* **13**, e0206678 (2018).

66. Lehner, P. et al. LUMOS—a sensitive and reliable optode system for measuring dissolved oxygen in the nanomolar range. *PLoS ONE* **10**, e0128125 (2015).
67. Sousa, F. L., Shavit-Grievink, L., Allen, J. F. & Martin, W. F. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol. Evol.* **5**, 200–216 (2013).
68. Gibson, J. A. E. et al. Geochemistry of ice-covered, meromictic lake A in the Canadian High Arctic. *Aquat. Geochem.* **8**, 97–119 (2002).
69. Green, W. J., Ferdelman, T. G. & Canfield, D. E. Metal dynamics in Lake Vanda (Wright Valley, Antarctica). *Chem. Geol.* **76**, 85–94 (1989).
70. Dickman, M. & Ouellet, M. Limnology of Garrow Lake, NWT, Canada. *Polar Rec. (Gr. Brit.)* **23**, 531–549 (1987).
71. Gallagher, J. B. in *Antarctic Nutrient Cycles and Food Webs* (eds. Siegfried, W. R. et al.) 234–237 (Springer, 1985).
72. Savvichev, A. S. et al. Microbial processes of the carbon and sulfur cycles in an ice-covered, iron-rich meromictic lake Svetloe (Arkhangelsk region, Russia). *Environ. Microbiol.* **19**, 659–672 (2017).
73. Tebo, B. M. Manganese (II) oxidation in the suboxic zone of the Black Sea. *Deep-Sea Res.* **38**, S883–S905 (1991).
74. Su, J. et al. CotA, a multicopper oxidase from *Bacillus pumilus* WH4, exhibits manganese-oxidase activity. *PLoS ONE* **8**, e60573 (2013).
75. Su, J. et al. Catalytic oxidation of manganese (II) by multicopper oxidase CueO and characterization of the biogenic Mn oxide. *Water Res.* **56**, 304–313 (2014).

Acknowledgements We thank the current members of the Bosak laboratory, the Simons Foundation Collaboration on the Origins of Life (no. 327126 to T.B. and no. 339603 to G.F.), FESD NSF project (no. 1338810 to T.B.) and NSF Integrated Earth Systems (no. 1615426 to G.F.

and T.B.). The NSF award number DMR-1419807 funded MIT Center for Material Science and Engineering (part of Materials Research Science and Engineering Center, NSF ECCS. award no. 1541959) funded the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI). The DOE Office of Science User Facility under contract no. DE-AC02-05CH11231 supports the Advanced Light Source and BI L12.3.2. M.P. and T.B. received funding from the John Templeton foundation, and M.P. was also supported by the Professor Amar G. Bose Research Grant Program (MIT).

Author contributions M.D. and T.B. conceived and designed the project. M.D. performed microbial enrichment experiments and analysed data. M.D. and N.T. performed synchrotron μ XRD diffraction and analysed data. M.D. and M.P. performed oxygen concentration measurement experiments and analysed data. V.K.-C. and S.R. analysed the composition of microbial communities and performed bioinformatics analyses. V.K.-C. and A.F.-S. conducted bioinformatics analyses of the co-cultures and pure culture genomes. G.F. wrote about the molecular evolution of photosynthetic groups and contributed Extended Data Fig. 6. N.B. contributed to the writing by commenting on Archaeal geology. M.D. and T.B. wrote the manuscript with input from M.P. and V.K.-C., N.B. and G.F. All co-authors reviewed and approved the final manuscript.

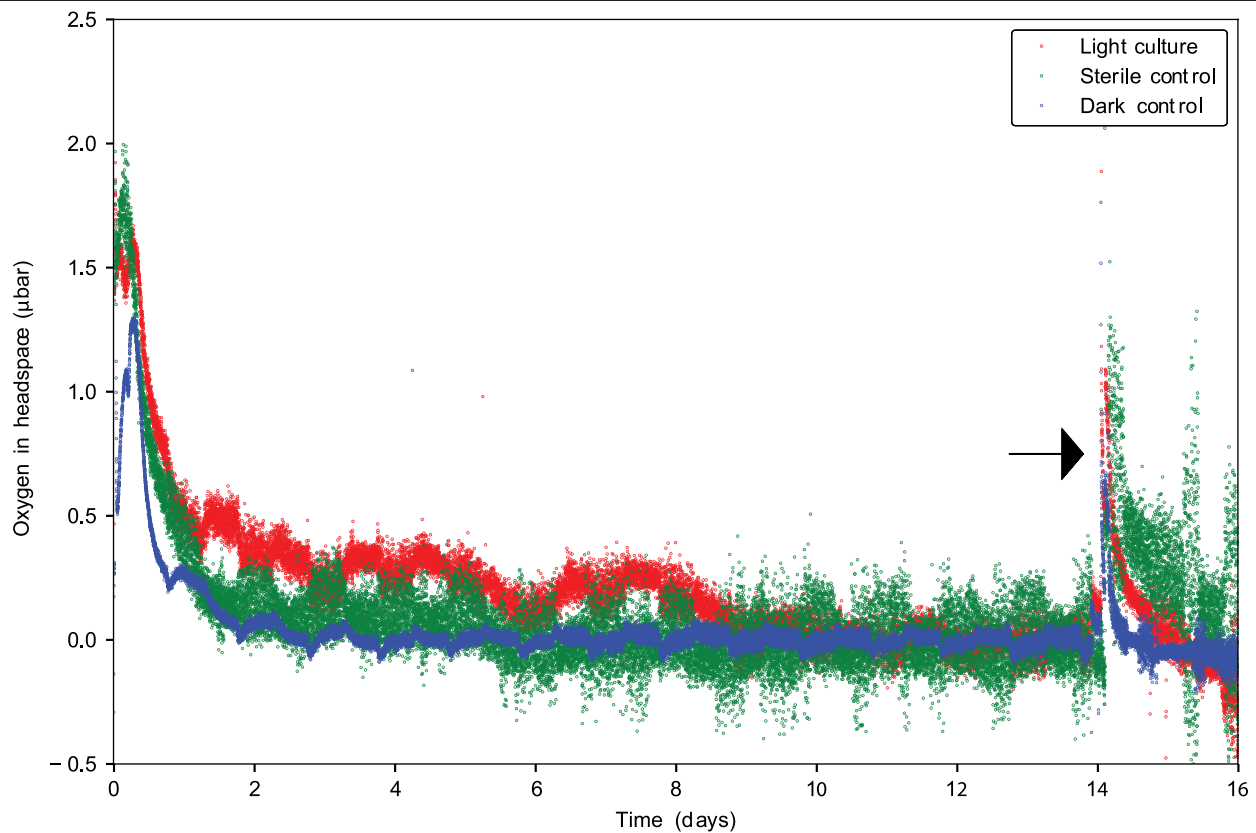
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1804-0>.

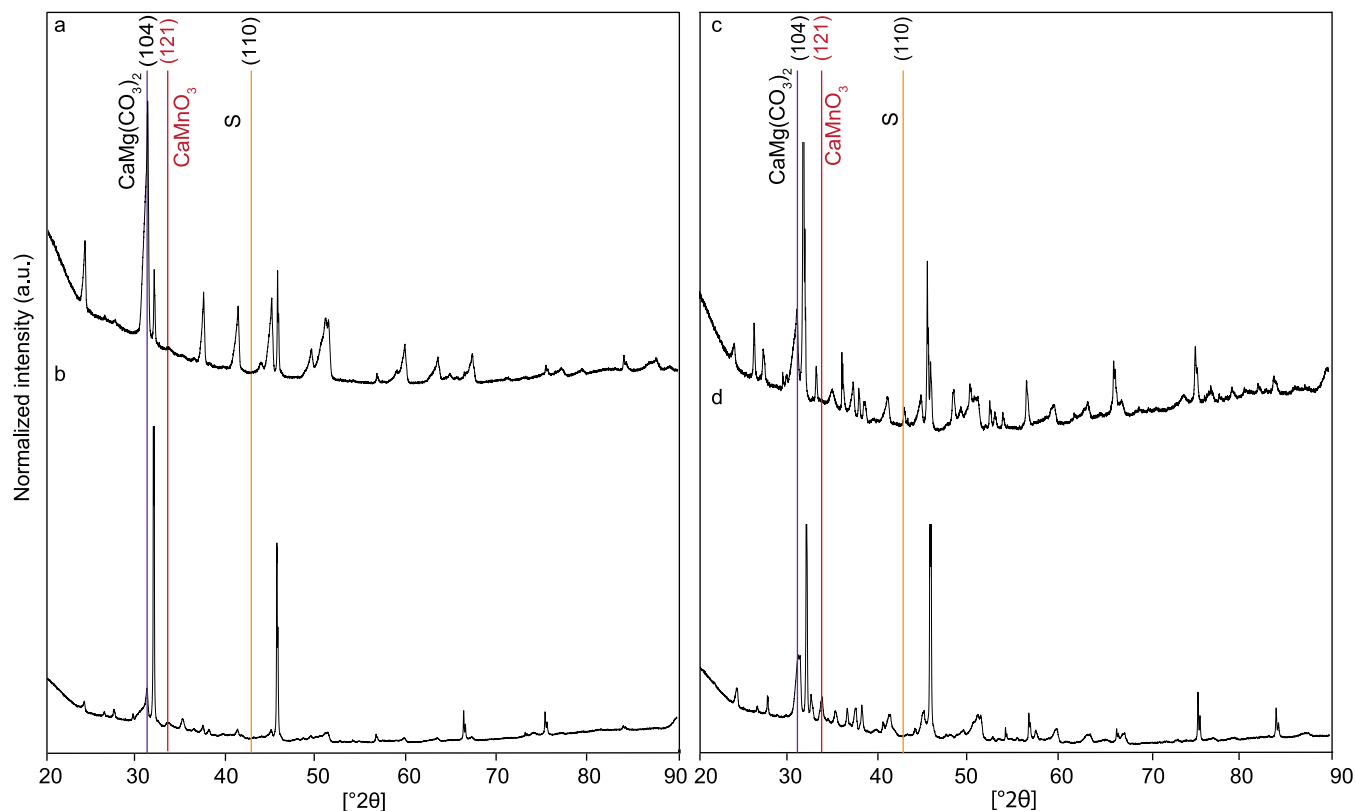
Correspondence and requests for materials should be addressed to M.D. or T.B.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



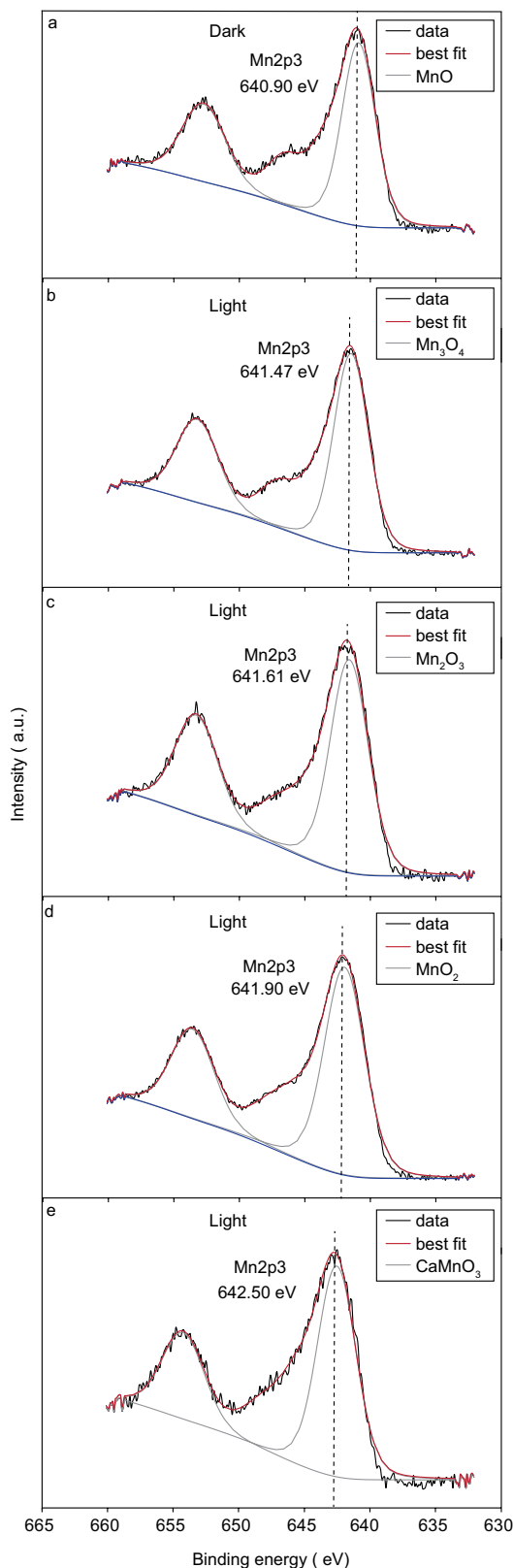
Extended Data Fig. 1 | Partial pressure of oxygen in the headspace of enrichment cultures and dark controls. Oxygen concentration (μatm) measured in the headspaces of 150-ml serum bottles that contained 100 ml of MFGI medium, 50 μM sulfide and 1 mM MnCl_2 . One inoculated culture was incubated in the light (red points) and another in the dark (blue points). The sterile control (green points) was incubated in the light. Individual points are measurements by the oxygen sensor taken every 48.2 s. To control for sensor

drift and recalibrate the zero point of the sensor, the bottles were flushed with oxygen-free N_2 on day 14 (black arrow) after the inoculation. The fluorescence reading value after the stabilization was set as zero. The diurnal oscillations in O_2 concentration reflect temperature changes induced by the proximity to the light bulb with a 12:12 h day:night cycle. Oxygen concentrations in all cultures were lower than 1 nM at all times after about 12 h and before the flushing on day 14. All data are representative of two independent measurements.

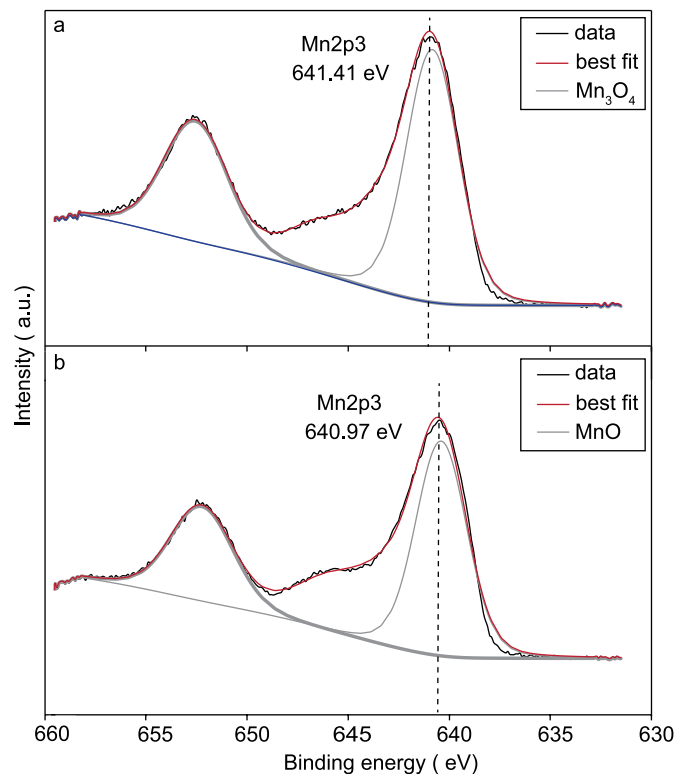


Extended Data Fig. 2 | XRD spectra of biofilm samples incubated in the light for two weeks. a–d. These biofilms were not treated to remove manganese oxides before inoculation. **a.** One millimolar Mn(II) and 0.05 mM Na₂S. **b.** Mn(II) at 0.1 mM, and 0.05 mM Na₂S. **c.** One millimolar Mn(II) and 1 mM Na₂S. **d.** One millimolar Mn(II) and 0.25 mM Na₂S. Purple line shows the highest intensity

peak at 2θ of 30.870° for the basal reflection of (104) plane of dolomite, CaMg(CO₃)₂. Red line shows the highest intensity peak at 2θ of 33.867° for the basal reflection of (121) plane of CaMnO₃. Orange line shows the highest intensity peak at 2θ of 42.845° for the basal reflection of (110) plane of S⁰. All data are representative of three independent measurements.



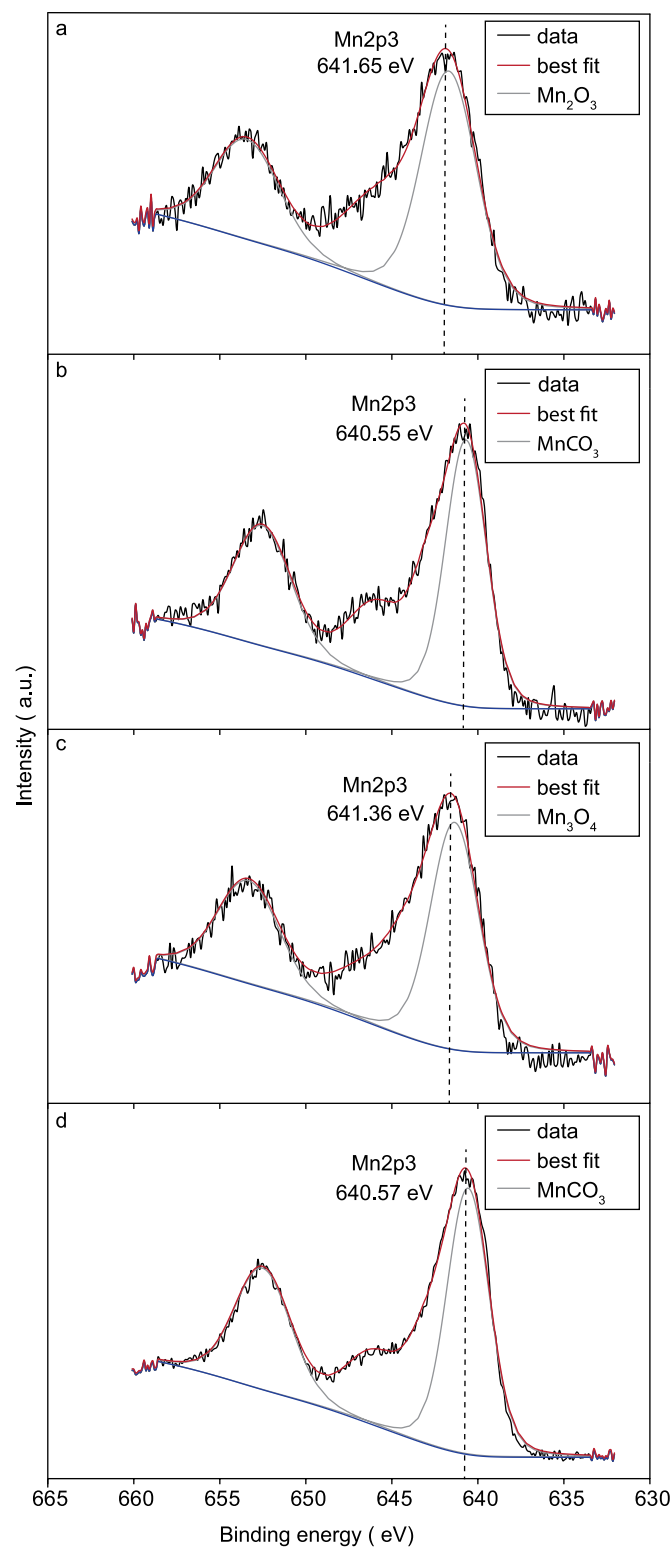
Extended Data Fig. 3 | XPS of the $2p$ spectral region of manganese in two-week-old microbial cultures. **a**, Biofilm incubated in the dark. The $Mn2p_{3/2}$ main peak of the sample fits the MnO standard at binding energy of 640.90 eV that corresponds to the redox state of Mn(II). **b**, Biofilm incubated in the light. The $Mn2p_{3/2}$ main peak of the sample fits the Mn_3O_4 standard at binding energy of 641.47 eV that corresponds to Mn(III) and Mn(II). **c**, Biofilm incubated in the light (a different region to that shown in **b** and **d**). The $Mn2p_{3/2}$ main peak of the sample fits the Mn_2O_3 standard at binding energy of 641.61 eV that corresponds to Mn(III). **d**, Biofilm incubated in the light (a different region to that shown in **b**, **c**). The $Mn2p_{3/2}$ main peak of the sample fits the MnO_2 standard at binding energy of 641.90 eV that corresponds to redox state of Mn(IV). **e**, Biofilm incubated in the light (a different region to that shown in **c**, **d**). The $Mn2p_{3/2}$ main peak of the sample fits the $CaMnO_3$ standard at binding energy of 642.50 eV that corresponds to Mn(IV). All data are representative of three independent measurements.



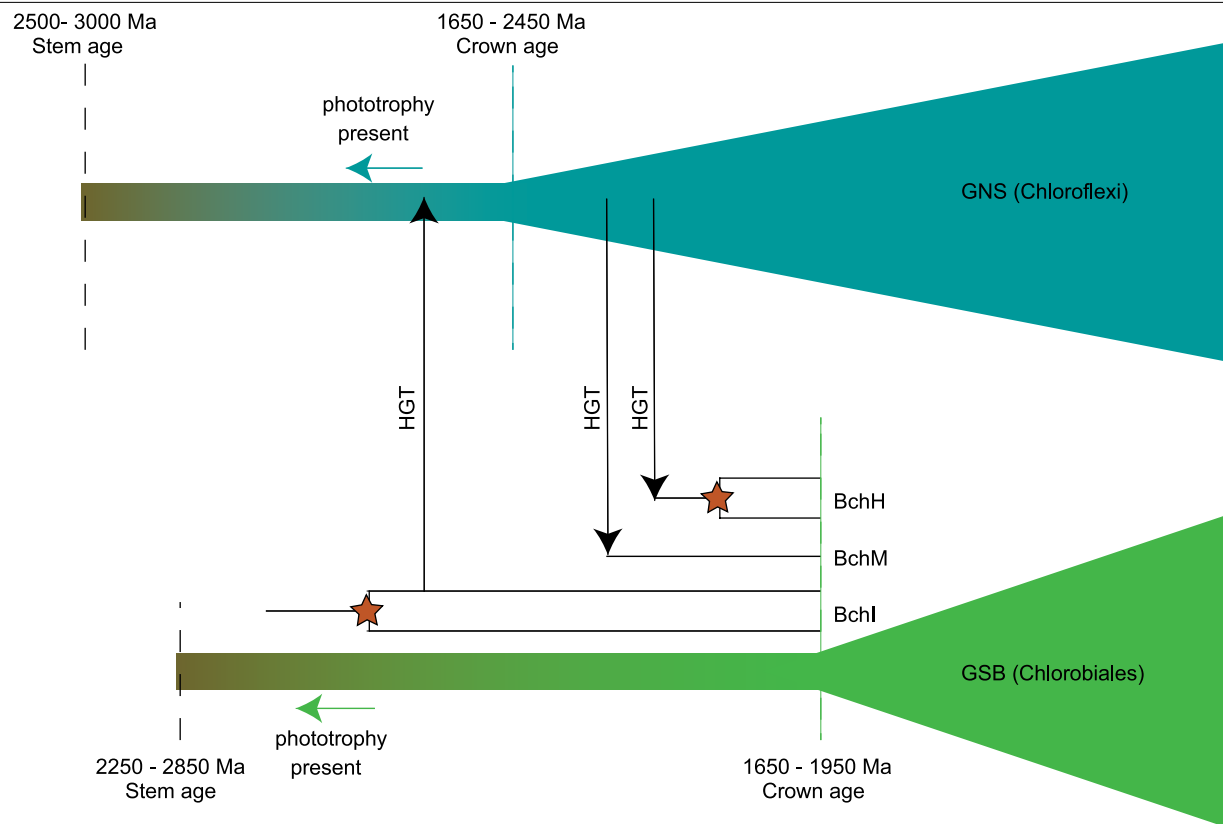
Extended Data Fig. 4 | Test of manganese-oxidizing activity in cell suspensions of photosynthetic cultures enriched under two conditions.

Results of XPS analysis of the 2p spectral region of manganese are shown.

a. Culture enriched on 1 mM Mn(II) and 0.05 mM H₂S (condition (1)). The Mn2p_{3/2} main peak of the sample fits Mn₃O₄ standard at binding energy of 641.41 eV. This corresponds to Mn(II) and Mn(III). **b.** Culture enriched on 1 mM H₂S (condition (2)). The Mn2p_{3/2} main peak of the sample fits MnO standard at binding energy of 640.97 eV and corresponds to the redox state Mn(II). A detailed experimental protocol is described in 'Interpretation of XPS spectra' in the Methods, and summarized in Extended Data Table 3. All data are representative of three independent measurements.



Extended Data Fig. 5 | Test of manganese-oxidizing activity in cell suspensions of pure cultures and co-cultures of *C. limicola*, *C. tepidum* and *G. lovleyi*. **a–d**, Results of XPS analysis of the 2p spectral region of manganese are shown. **a**, *C. limicola*, *C. tepidum* and *G. lovleyi*. The Mn2p_{3/2} main peak of the sample fits Mn₂O₃ standard at binding energy of 641.65 eV. This corresponds to a valence state of Mn(III). **b**, *C. limicola* and *C. tepidum*. The Mn2p_{3/2} main peak of the sample fits MnCO₃ standard at binding energy of 640.55 eV and corresponds to the redox state Mn(II). **c**, *C. limicola* and *G. lovleyi*. The Mn2p_{3/2} main peak of the sample fits Mn₃O₄ standard at binding energy of 641.36 eV. **d**, *C. tepidum* and *G. lovleyi*. The Mn2p_{3/2} main peak of the sample fits MnCO₃ standard at binding energy of 640.57 eV. A detailed experimental protocol is described in 'Probing the redox state of manganese in co-cultures' in the Methods. All co-cultures were grown with 1 mM Mn(II) and 0.05 mM H₂S for 2 weeks in the light. All data are representative of three independent measurements.



Extended Data Fig. 6 | Reticulate history of bacteriochlorophyll-biosynthesis genes supports a long history of phototrophy in the Chlorobiales stem lineage. Horizontal gene transfers and gene duplications of

bacteriochlorophyll genes were taken from a previous publication⁶⁷. Age estimates for crown Chlorobi and GNS groups were taken from a previous publication²⁶.

Extended Data Table 1 | Aquatic environments with H₂S and manganese in the photic zone

Water body	H ₂ S (μM)	Mn(II) (μM)	Photic Zone	Reference
Lake A	230	140	Green Sulfur Bacteria	68
Lake Vanda	240	120	Unclear	69
Garrow Lake	20	18	Green Sulfur Bacteria	70
Sombre Lake	1.2	68	Green Sulfur Bacteria	71
Svetloe Lake	2	60	Green Sulfur Bacteria	72
Black Sea	2	8.4	Green Sulfur Bacteria	73
Green Lake	20-30	50-60	Green Sulfur Bacteria	8

Data are from previous publications^{8,68-73}.

Extended Data Table 2 | Shake tube and transfer procedures used to obtain enrichments from conditions 1 and 2

First round shake tube	Incubation period, (days)	Colonies	Transfer into liquid medium	Incubation period (days)	Growth ¹	Enrichment	
20 μ M Na ₂ S, 1 mM MnCl ₂	30	Black and dark brown	Brown colony; 0.02 mM Na ₂ S, 1 mM MnCl ₂	30	+	Condition 1	
1 mM Na ₂ S	30	Black, white and dark brown	Brown colony; 1 mM Na ₂ S	30	+++	Condition 2	
Second round shake tube	Incubation period (days)	Colonies	Transfer into liquid medium	Incubation period (days)	Growth ¹	Enrichment	Microbes present in enrichment culture ²
0.02 mM Na ₂ S, 1 mM MnCl ₂	30	Brown, dark brown	Brown colony; 0.02 mM Na ₂ S, 1 mM MnCl ₂	30	+	Condition 1	<i>Chlorobium limicola</i> , <i>Geobacter lovleyi</i> , others ³
1 mM Na ₂ S	30	Brown, dark brown	Brown colony; 1 mM Na ₂ S	30	+++	Condition 2	<i>Chlorobium limicola</i> , <i>Desulfomicrobium sp.</i>

¹Growth of colonies transferred from shake to a liquid medium of the same chemical composition. +, low microbial growth; +++, high microbial growth.

²Microbial composition determined by metagenomic sequencing.

³*A. equifetale*, *Alistipes sp.* HGB5, and *C. oshimai*.

Extended Data Table 3 | Summary of the methods used to examine the redox state of manganese in enrichment cultures

Enrichment condition	Transfer condition	Transfer MFGL medium composition	Incubation time (days)	Growth ¹	Mn oxidation activity ²
Condition 1	Cell suspension	0.02 mM Na ₂ S, 1mM MnCl ₂	3	-	+
Condition 2	Cell suspension	0.02 mM Na ₂ S, 1 mM MnCl ₂	3	-	-
Condition 1	5 % inoculum	0.02 mM Na ₂ S, 1mM MnCl ₂	7	-	-
Condition 2	5 % inoculum	0.02 mM Na ₂ S, 1mM MnCl ₂	7	+	-

¹+ or – indicates visible growth (+) of the co-culture, or not (–)
²Manganese oxidation in the enrichment cultures was examined using XPS; + or –, presence (+) or absence (–) of Mn(II)-oxidizing activity in the microbial co-culture.

Extended Data Table 4 | Mn(II)-oxidation genes with confirmed function compared using BLASTp v. 2.6.0+ against the genome of *C. limicola* SR-12

Locus ID	Gene Annotation	Gene Name	Organism	Ref.	% identity	e. value	bit score
PputGB1_3353	animal heme peroxidase	<i>mopA</i>	<i>Pseudomonas putida</i>	18	46	1.0E-08	57.4
WP_007817484	animal heme peroxidase	<i>ahpL</i>	<i>Roseobacter</i> sp. AzwK-3b	19	40	7.0E-06	48.5
WP_009209951	animal heme peroxidase	<i>mopA</i>	<i>Aurantimonas manganoxydans</i>	51	47	1.0E-07	53.5
WP_006837219	multi-copper oxidase	<i>mxnG</i>	<i>Bacillus</i> sp. strain SG-1	17	no hits		
WP_076798083	multi-copper oxidase / billirubin oxidase	<i>boxA</i>	<i>Arthrobacter</i> sp. QXT-31	32	no hits		
AFL56752	multi-copper oxidase, type 2	<i>cotA</i>	<i>Bacillus pumilus</i> WH4	74	no hits		
CAJ19378	multi-copper oxidase	<i>moxA</i>	<i>Pedomicrobium</i> sp. ACM 3067	50	no hits		
NP_745328	multi-copper oxidase / billirubin oxidase	<i>mcoA</i>	<i>Pseudomonas putida</i>	18	no hits		
EG12318	multi-copper oxidase	<i>cueO</i>	<i>Escherichia coli</i>	75	no hits		

All animal haem peroxidases also hit the haemolysin-type calcium-binding region in *C. limicola* ($e = 3 \times 10^{-21}$, bit score = 98.6). Data are from previous publications^{17-19,32,50,51,74,75}.

Corresponding author(s):

M. Daye et al., 2018-06-07850D

☐ Initial submission ☐ Revised version ☒ Final submission

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software has been used to collect data.

Data analysis

Microsoft excel has been used to analyze, plot and perform statistical analyses of the data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data is deposited on Figshare and made available through weblinks.

The data will be published on figshare and will be made available for public through Digital Object Identifier (DOI) weblinks after publication of the manuscript.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size (s) calculation was performed. For each experiment we do test for different chemical conditions and for each variable we have triplicates (n=3). Each experiment can have at least 4 different variables (v=3) which is equivalent: $s=n*v=12$
Data exclusions	No data exclusion performed in this study excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Replication	For each analysis, triplicates were performed. Each experiment was at least 3 times repeated to ensure reproducibility. Each experiment was repeated independently two five times to ensure similar results. Describe why.
Randomization	This is not relevant to the study. We are examining one type of microbial biofilm. Describe how covariates were controlled OR if this is not relevant to your study, explain why.
Blinding	Not relevant to the study since we are studying the manganese anaerobic oxidation in one type of photosynthetic microbial biofilm. possible, describe why blinding was not relevant to your study.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Randomization	If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
-------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.
Did the study involve field work?	<input type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access and import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials	Describe any restrictions on the availability of unique materials OR confirm that all unique materials used are readily available from the authors or from standard commercial sources (and specify these sources).
----------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Antibodies

Antibodies used	Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.
-----------------	------------------------------------------------------------------------------------------------------------------------------

Validation

Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

State the source of each cell line used.

Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session
(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.
Instrument	Identify the instrument used for data collection, specifying make and model number.
Software	Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.
Cell population abundance	Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.
Gating strategy	Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.
<input type="checkbox"/> Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.	

Magnetic resonance imaging

Experimental design

Design type	Indicate task or resting state; event-related or block design.
Design specifications	Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.
Behavioral performance measures	State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)	Specify: functional, structural, diffusion, perfusion.
Field strength	Specify in Tesla
Sequence & imaging parameters	Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.
Area of acquisition	State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See [Eklund et al. 2016](#))

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a | Involved in the study

- ☐ ☐ Functional and/or effective connectivity
☐ ☐ Graph analysis
☐ ☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.



Structure and drug resistance of the *Plasmodium falciparum* transporter PfCRT

<https://doi.org/10.1038/s41586-019-1795-x>

Received: 12 March 2019

Accepted: 6 November 2019

Published online: 27 November 2019

Jonathan Kim^{1,13}, Yong Zi Tan^{1,2,13}, Kathryn J. Wicht³, Satchal K. Erramilli⁴, Satish K. Dhingra³, John Okombo³, Jeremie Vendome⁵, Laura M. Hagenah³, Sabrina I. Giacometti¹, Audrey L. Warren⁶, Kamil Nosol⁴, Paul D. Roepe^{7,8}, Clinton S. Potter^{2,9}, Bridget Carragher^{2,9}, Anthony A. Kossiakoff⁴, Matthias Quick^{6,10,11*}, David A. Fidock^{3,12*} & Filippo Mancia^{1*}

The emergence and spread of drug-resistant *Plasmodium falciparum* impedes global efforts to control and eliminate malaria. For decades, treatment of malaria has relied on chloroquine (CQ), a safe and affordable 4-aminoquinoline that was highly effective against intra-erythrocytic asexual blood-stage parasites, until resistance arose in Southeast Asia and South America and spread worldwide¹. Clinical resistance to the chemically related current first-line combination drug piperazine (PPQ) has now emerged regionally, reducing its efficacy². Resistance to CQ and PPQ has been associated with distinct sets of point mutations in the *P. falciparum* CQ-resistance transporter PfCRT, a 49-kDa member of the drug/metabolite transporter superfamily that traverses the membrane of the acidic digestive vacuole of the parasite^{3–9}. Here we present the structure, at 3.2 Å resolution, of the PfCRT isoform of CQ-resistant, PPQ-sensitive South American 7G8 parasites, using single-particle cryo-electron microscopy and antigen-binding fragment technology. Mutations that contribute to CQ and PPQ resistance localize primarily to moderately conserved sites on distinct helices that line a central negatively charged cavity, indicating that this cavity is the principal site of interaction with the positively charged CQ and PPQ. Binding and transport studies reveal that the 7G8 isoform binds both drugs with comparable affinities, and that these drugs are mutually competitive. The 7G8 isoform transports CQ in a membrane potential- and pH-dependent manner, consistent with an active efflux mechanism that drives CQ resistance⁵, but does not transport PPQ. Functional studies on the newly emerging PfCRT F145I and C350R mutations, associated with decreased PPQ susceptibility in Asia and South America, respectively^{6,9}, reveal their ability to mediate PPQ transport in 7G8 variant proteins and to confer resistance in gene-edited parasites. Structural, functional and in silico analyses suggest that distinct mechanistic features mediate the resistance to CQ and PPQ in PfCRT variants. These data provide atomic-level insights into the molecular mechanism of this key mediator of antimalarial treatment failures.

Recent reductions in the global malaria burden have stalled, and the situation in Southeast Asia is exacerbated by the emerging resistance to artemisinin-based combination therapies, including the bis-4-aminoquinoline partner drug PPQ¹⁰. A major determinant of drug resistance in malaria parasites is the *P. falciparum* transporter PfCRT⁵. Amino acid substitutions in this protein comprise haplotypes that originated independently decades ago in several regions that were subject to

intense drug pressure with CQ, the former gold-standard antimalarial drug. These include the five-amino acid 7G8 variant that dominates in South America and the Western Pacific, and the eight-amino acid Dd2 variant that is prevalent in Southeast Asia (Extended Data Fig. 1). Recent extensive use of PPQ in Cambodia is suspected to have driven the rapid emergence of novel mutations in *pfcr*, giving rise to the Dd2 isoform^{6–8}. These mutations have become widespread across Southeast Asia, where

¹Department of Physiology and Cellular Biophysics, Columbia University Irving Medical Center, New York, NY, USA. ²National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA. ³Department of Microbiology and Immunology, Columbia University Irving Medical Center, New York, NY, USA. ⁴Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA. ⁵Schrödinger, New York, NY, USA. ⁶Department of Psychiatry, Columbia University Irving Medical Center, New York, NY, USA. ⁷Department of Chemistry, Georgetown University, Washington, DC, USA. ⁸Department of Biochemistry and Cellular and Molecular Biology, Georgetown University, Washington, DC, USA. ⁹Department of Biochemistry and Molecular Biophysics, Columbia University Irving Medical Center, New York, NY, USA. ¹⁰Center for Molecular Recognition, Columbia University Irving Medical Center, New York, NY, USA. ¹¹Division of Molecular Therapeutics, New York State Psychiatric Institute, New York, NY, USA. ¹²Division of Infectious Diseases, Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA. ¹³These authors contributed equally: Jonathan Kim, Yong Zi Tan. *e-mail: mq2102@cumc.columbia.edu; df2260@cumc.columbia.edu; fm123@cumc.columbia.edu

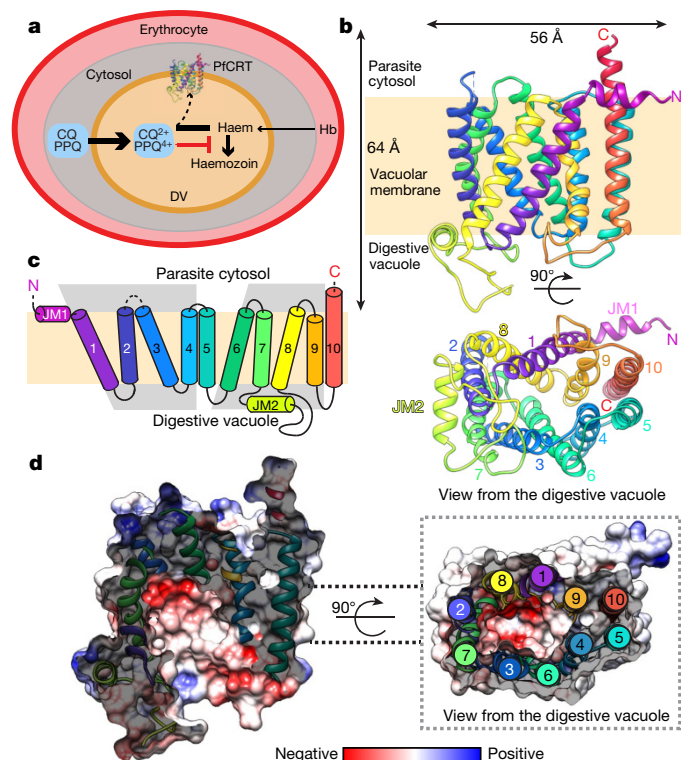


Fig. 1 | Single-particle cryo-EM structure of PfCRT 7G8. **a**, PfCRT (PlasmoDB PF3D7_0709000) is localized within the membrane of the digestive vacuole (DV) of the *P. falciparum* intra-erythrocytic parasite, in which imported host haemoglobin (Hb) is catabolized and toxic free haem is released. CQ and PPQ are believed to concentrate in the digestive vacuole as protonated species (CQ^{2+} and PPQ^{4+}) that bind to haem and prevent its incorporation into non-toxic haemozoin⁵. In CQ-R parasites, PfCRT is thought to efflux CQ out of the digestive vacuole into the cytosol away from its haem target. **b**, The 3.2 Å cryo-EM structure of PfCRT 7G8, with the 10 TM helices coloured in rainbow. The N and C termini are labelled. Bottom, the structure of PfCRT 7G8 rotated 90° compared to the top, as viewed from the digestive vacuole side. TM and JM helices are numbered. **c**, Topology of PfCRT highlighting the inverted antiparallel repeats of TM1–TM4 and TM6–TM9 (shown as grey trapezoids). Disordered regions are shown as dotted lines. TM1–TM10 and JM1 and JM2 are labelled. TM1–TM4 and TM6–TM9 surround the central cavity. **d**, Surface representation of the electrostatic potential of the central cavity with red and blue indicating negatively and positively charged residues, respectively. On the right, a central slice through the structure (dotted lines) as an inset shows the arrangement of TM helices, labelled from N terminus to C terminus.

they are associated with dihydroartemisinin–PPQ treatment failures that now occur, on average, in 50% of cases in the region and reach 87% of cases in northeastern Thailand^{2,7}. CQ and PPQ are thought to act by accumulating in the acidic digestive vacuole of the intra-erythrocytic parasite as protonated species (CQ^{2+} and PPQ^{4+}). These drugs bind to toxic Fe^{3+} -haem, which is released from proteolysed host haemoglobin, and inhibit haem incorporation into chemically inert haemozoin^{11,12}. Resistance to CQ has been attributed to mutant PfCRT-mediated drug efflux out of the digestive vacuole^{4,13–16} (Fig. 1a), whereas the mechanism of PPQ resistance has remained unclear.

We used single-particle cryo-electron microscopy (cryo-EM) to determine the structure of the 49-kDa PfCRT 7G8 isoform, which was identified to be the most suitable candidate from a panel of 12 sequence variants and/or orthologues expressed in HEK293 cells. To overcome current cryo-EM size limitations¹⁷, we screened a synthetic phage display library¹⁸ to select for recombinant PfCRT-specific antigen-binding fragments (Fabs, around 50 kDa). PfCRT complexes were obtained with several candidates, including the Fab CTC that yielded high-quality

two-dimensional class averages obtained by negative-stain electron microscopy (Extended Data Fig. 2).

We solved the structure of the nanodisc-incorporated PfCRT 7G8–Fab CTC complex to a nominal resolution of 3.3 Å, with the PfCRT portion resolved to 3.2 Å (Extended Data Table 1 and Extended Data Fig. 3). We built a de novo model of PfCRT comprising residues 47–113 and 123–405 of the 424-amino acid protein (Fig. 1b, c). PfCRT appears to be monomeric and consists of 10 transmembrane (TM) helices (Fig. 1b). Two juxtamembrane (JM) helices (JM1 and JM2), parallel to and on opposite sides of the digestive vacuole membrane, are located at the cytosolic N terminus and between TM7 and TM8, respectively (Fig. 1b, c and Extended Data Fig. 4). The 10 TM helices are arranged as five helical pairs that form two-helix hairpins with an inverted antiparallel topology (Extended Data Fig. 5a). TM1–TM4 and TM6–TM9 form a central cavity of around 3,300 Å³ that is wider on the digestive vacuole side and closes approximately halfway into the membrane, suggesting that the structure is in an open-to-digestive-vacuole/inward-open conformation (Fig. 1b, d). The pseudo-symmetrical arrangement of the TM helices also enabled us to model the alternative open-to-cytosol conformation of TM1–TM10 (Extended Data Fig. 5b).

Structural conservation between PfCRT 7G8 and other members of the drug/metabolite transporter (DMT) superfamily with known structure is high in the TM helices, with a root mean square deviation of 2.8–4.0 Å between backbone atoms (Extended Data Fig. 5c, d). The central cavity in the structurally conserved DMT proteins constitutes the known ligand-binding site (Extended Data Fig. 5d), indicating that this cavity is probably the equivalent site in PfCRT. The net charge in the PfCRT 7G8 cavity is negative (Fig. 1d and Extended Data Fig. 5d), influenced mainly by D137, D326 and D329 (Extended Data Fig. 6a), suggesting that the putative substrates of PfCRT 7G8 are probably positively charged. Consistent with this proposal, the positively charged R111 residue in the variable loop region of the Fab CTC protrudes into the PfCRT cavity (Extended Data Fig. 6a). Interactions between Fab R111 and PfCRT D326 and D329 presumably help to stabilize the PfCRT 7G8–Fab complex. We also identified density in a separate hydrophobic cleft—which is delineated by JM1, TM1, TM9 and TM10—that is unique to PfCRT and near the cytosolic side of the digestive vacuole membrane. We tentatively assigned this to cholesteryl hemisuccinate, which was present during the purification (Extended Data Fig. 6b–d).

The PfCRT central cavity, with a maximum diameter of approximately 25 Å, is in principle able to accommodate CQ or PPQ (which have maximum lengths of approximately 14 and 21 Å, respectively; Extended Data Fig. 6e). All CQ-resistant (CQ-R) isoforms, irrespective of geographical origin, share a common and essential K76T mutation^{13,19}, which our structure places directly in the lining of the cavity (Fig. 2). Compared with the canonical wild-type CQ-sensitive (CQ-S) 3D7 isoform, the CQ-R 7G8 isoform has four additional mutations (C72S, A220S, N326D and I356L) that also line the cavity (Fig. 2 and Extended Data Table 2a). Similarly, the CQ-R Dd2 isoform has six of its eight mutations (M74I, N75E, K76T, A220S, N326S and I356T) located within the cavity, and two additional mutations (Q271E and R371I) are located on the loops at the cavity entrance on the digestive vacuole side (Fig. 2). Distinct amino acid substitutions associated with PPQ-R also line the cavity, including the C350R mutation that arose in the 7G8 isoform, and H97Y, F145I and G353V, as well as the recently described mutations T93S and I218F, which each emerged separately in the Dd2 isoform^{2,6–9} (Fig. 2 and Extended Data Table 2a). M343L, which confers low-grade PPQ resistance in the Dd2 isoform⁶, lies deeper into the central cavity, close to the parasite cytosol (Fig. 2).

CQ resistance-associated mutations are confined mostly to a single region that includes TM1, TM6, TM8 and TM9, in contrast to the PPQ mutations that are positioned primarily in two distinct regions that involve TM2, TM3, TM6 and TM9 (Extended Data Fig. 6f). All PPQ-R point mutations arose individually in the CQ-R Dd2 or 7G8 isoforms, indicating that the interaction with the larger and more positively

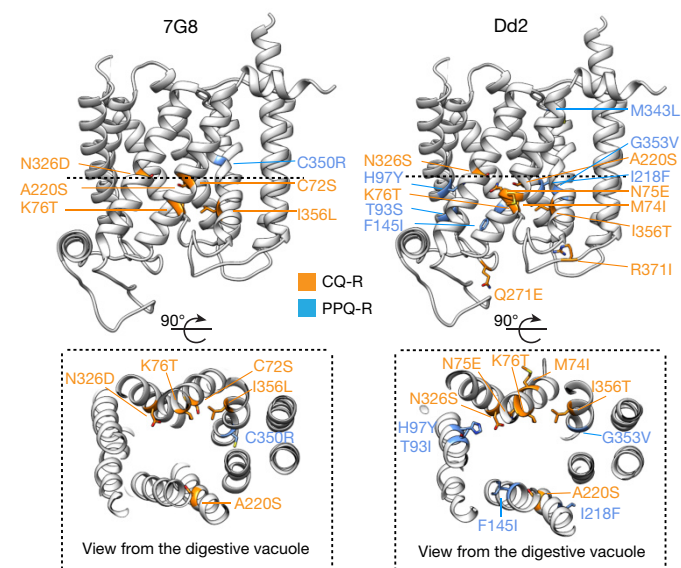


Fig. 2 | Mapping of drug-resistance mutations onto the PfCRT structure. Residues known to contribute to resistance to CQ and PPQ (Extended Data Fig. 1 and Extended Data Table 2) in the 7G8 and Dd2 isoforms are mapped onto the 7G8 structure. Mutations have their side chains rendered as sticks and are coloured based on their associated resistance profiles. The remaining structures are rendered in cartoon and coloured in grey. Views are shown vertically (digestive vacuole to the bottom) and from the digestive vacuole side.

charged PPQ⁴⁺ requires additional modifications to the central cavity. Potentially, the physical separation between the sets of mutated helices in the PPQ-R isoforms might correlate with the double 4-aminoquinoline ring structure of PPQ, in contrast to the pattern observed for CQ, which contains a single 4-aminoquinoline ring (Extended Data Fig. 6e, f). No mutations in PfCRT associated with resistance to either drug mapped to residues that were fully conserved across our set of apicomplexan CRT orthologues (Extended Data Figs. 6f, 7 and Extended Data Table 2a), suggesting that there are structural and/or functional constraints. Resistance-conferring mutations map to moderately conserved sites in select helices in residues that line the cavity, thus providing a road map of sites for future PPQ resistance surveillance. For example, residues with these properties on TM2 are at positions 90, 93, 94, 97, 98 and 101, among which T93S, H97Y and C101F have each been associated with PPQ resistance^{6–8,12}. By contrast, mutations previously found in Southeast Asia that have not recently expanded in areas of PPQ treatment include L196P, N295I, A366T and G367C, which do not meet the above criteria^{6,7}.

To translate structural observations into transporter function, we first used purified protein in scintillation proximity-based assays to measure drug binding²⁰. Results with PfCRT 7G8 reconstituted in nanodiscs showed binding to [³H]CQ and [³H]PPQ, at comparable levels (Fig. 3a). Isotopic dilution assays yielded mean half-maximal binding constants (EC₅₀) of approximately 300 nM for CQ, similar to previous studies with Dd2 PfCRT²¹, and around 190 nM for PPQ (Fig. 3b). Each drug competitively inhibited binding of the other with nearly identical mean half-maximal inhibition constants (IC₅₀ values) of approximately 170 nM (Fig. 3c).

Binding of these two drugs was partially inhibited by verapamil (Fig. 3d), an agent that is known to reverse CQ resistance²². Near-complete inhibition was observed with amodiaquine, a related 4-aminoquinoline first-line partner drug against which PfCRT 7G8 mediates cross-resistance²³. Control assays found no inhibition of [³H]CQ or [³H]PPQ binding with the chemically distinct antimalarial drugs lumefantrine and atovaquone, which have separate modes of action⁵.

We also observed inhibition of [³H]CQ and [³H]PPQ binding with excess arginine, consistent with the observed interaction between R111 of the Fab CTC and the PfCRT cavity (Extended Data Fig. 6a), and previous [³H]Arg transport studies²⁴. Control experiments revealed no inhibition of drug binding with the neutral amino acid leucine (Fig. 3d). Results of the scintillation proximity-based assays showed binding of [³H]Arg to PfCRT 7G8 with an EC₅₀ of approximately 400 μM (Extended Data Fig. 8a, b). This binding was effectively competed by both CQ and PPQ, as well as by amodiaquine and non-radiolabelled arginine or lysine, but not by leucine (Extended Data Fig. 8c). Binding of [³H]Arg, [³H]CQ and [³H]PPQ was inhibited by the Fab CTC in a concentration-dependent manner (Fig. 3e), suggesting that Arg, CQ, PPQ and the Fab overlap in their binding in the PfCRT cavity.

We next showed that active uptake of [³H]CQ and [³H]Arg required an inwardly directed pH gradient (pH 5.5–7.5, corresponding to the digestive vacuole pH gradient of the parasite²⁵), and a membrane potential ΔΨ (Extended Data Fig. 8d, e). Here, ΔΨ was created by using the K⁺-specific ionophore valinomycin to generate an outward-directed K⁺ diffusion gradient from the proteoliposomes. Under these conditions, uptake of [³H]CQ and [³H]Arg was maximal within 1 min (Fig. 3f). Kinetic characterization of the concentration-dependent uptake of [³H]CQ and [³H]Arg yielded Michaelis–Menten constants (K_m) of 0.8 μM and 1.3 μM and maximum velocities of transport (V_{max}) of approximately 90 nmol mg^{−1} min^{−1} and 190 nmol mg^{−1} min^{−1}, respectively (Fig. 3f). Assuming full activity of PfCRT 7G8 reconstituted in proteoliposomes, these V_{max} values reflect catalytic turnover numbers (k_{cat}) of around 0.2 s^{−1} and about 0.4 s^{−1}, respectively; values that are characteristic for secondary active transporters rather than for channels²⁶. Collectively, these results indicate that proton-motive force-dependent CQ efflux is the mechanism that underlies mutant PfCRT 7G8-mediated resistance to CQ, in agreement with previous reports^{4,13–15,22}. Under our experimental conditions, no detectable [³H]PPQ uptake was observed with PfCRT 7G8 in proteoliposomes, as expected for this PPQ-sensitive (PPQ-S) isoform (Fig. 3f). We also observed partial-to-strong inhibition of [³H]CQ and [³H]Arg uptake by verapamil and 4-aminoquinolines, further implicating a crucial role for the central cavity in substrate recognition (Fig. 3g).

Our structural analysis reveals four cysteines in the long loop between TM7 and TM8, located on the digestive vacuole side near the entrance to the cavity. These cysteines, which appear to form the disulphide bridges Cys289–Cys312 and Cys301–Cys309 (located next to JM2 and at the tip of the loop, respectively), are completely conserved (Extended Data Figs. 7, 9a), suggesting that these cysteines might have a potential redox-dependent mechanistic role. Of note, *P. falciparum* intra-erythrocytic parasites are subjected to oxidative stress during parasite-mediated haemoglobin proteolysis and Fe²⁺-haem oxidation⁵. We expressed and purified 7G8 variants that expressed either a C301A or C289A mutation. We observed equivalent [³H]CQ binding for these variants and 7G8, which was unaffected by increasing concentrations of the reducing agent β-mercaptoethanol (Extended Data Fig. 9b). [³H]CQ uptake was also similar between these three isoforms in the absence of a reducing agent. Increasing the concentration of β-mercaptoethanol led to reduced uptake of CQ by all three tested isoforms (Extended Data Fig. 9c). The reduced dose dependency of the 7G8 + C301A isoform in respect to 7G8 and 7G8 + C289A indicates that the Cys289–Cys312 pair is less sensitive to reducing conditions that appear to regulate CQ transport.

To relate PfCRT structure and function to resistance mechanisms, we leveraged mutations in Dd2 and 7G8 isolates that were implicated in differential CQ and PPQ susceptibilities. We chose the F145I mutation, which mediates high-grade resistance to PPQ in Dd2 parasites and which is associated with a substantially increased risk of failure of the dihydroartemisinin–PPQ treatment in Southeast Asia^{2,6,7,27}. We hypothesized that this mutation might also lead to PPQ resistance in the 7G8 (South American and Western Pacific) background. We also selected

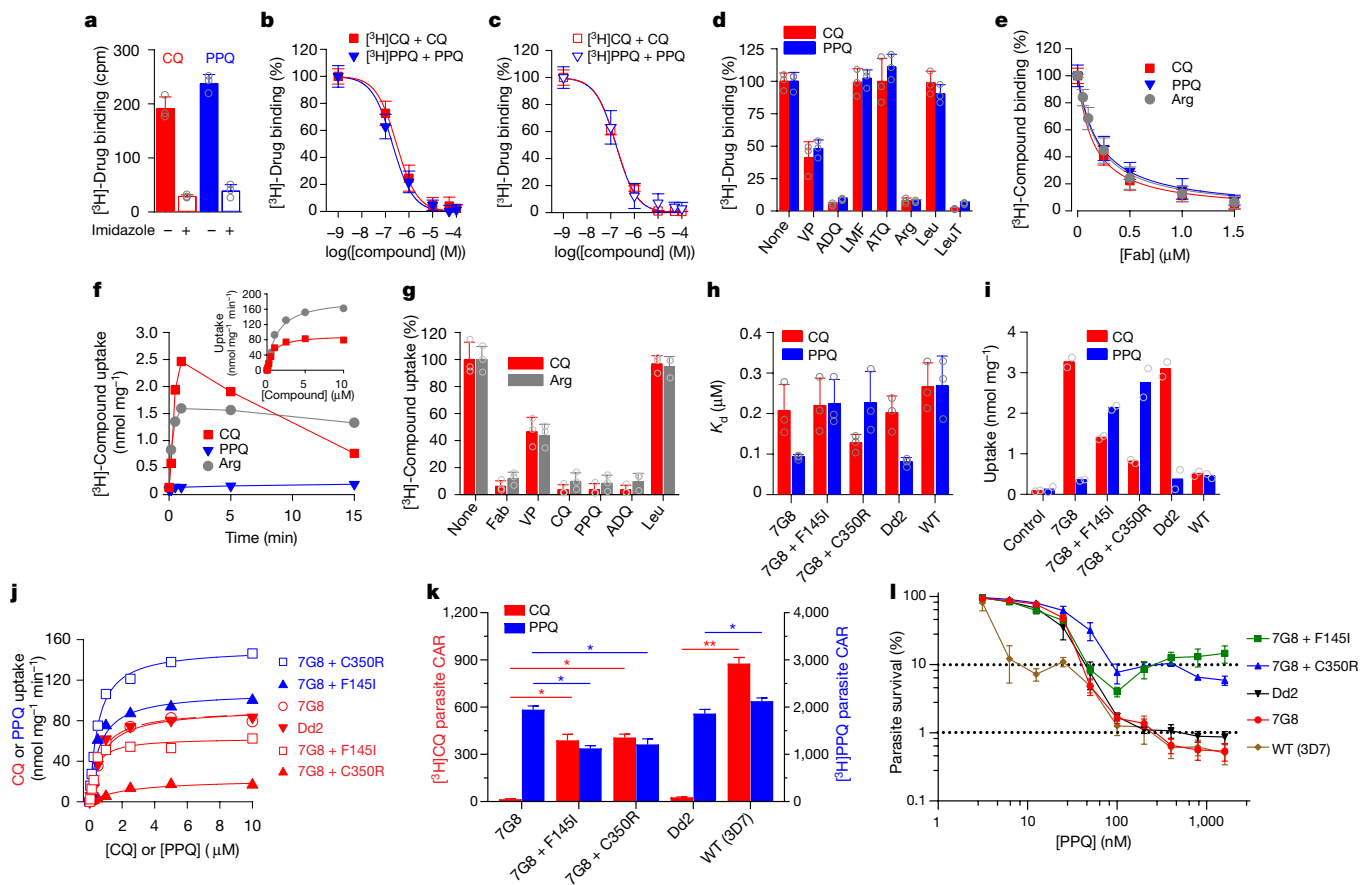


Fig. 3 | Functional characterization of PfCRT isoforms. **a**, Binding of 370 nM [3 H]CQ (red) or 75 nM [3 H]PPQ (blue) to PfCRT 7G8 with or without imidazole. cpm, counts per minute. Data are mean \pm s.e.m. of $n=3$ independent experiments; grey symbols are the data mean of technical triplicates of each independent experiment to show the data distribution. **b**, Isotopic dilution of [3 H]CQ and [3 H]PPQ revealed $\log(\text{EC}_{50}) = -6.53 \pm 0.04$ and $\log(\text{EC}_{50}) = -6.70 \pm 0.05$ (corresponding to means of 297 nM for CQ and 190 nM for PPQ), respectively. Data are mean \pm s.e.m. of $n=3$ independent experiments performed as technical triplicates. **c**, Competition of [3 H]CQ and [3 H]PPQ binding with non-radiolabelled PPQ and CQ, respectively, revealed $\log(\text{IC}_{50}) = -6.77 \pm 0.03$ and $\log(\text{IC}_{50}) = -6.78 \pm 0.06$ (corresponding to means of 171 nM and 167 nM), respectively. Data are mean \pm s.e.m. of $n=3$ independent experiments performed as technical triplicates. **d**, Specific binding of [3 H]CQ or [3 H]PPQ performed in the absence (–) or presence of 10 μM verapamil (VP), 1 μM amodiaquine (ADQ), 10 μM lumefantrine (LMF) or atovaquone (ATQ), or 1 mM Arg or Leu. Nanodiscs containing LeuT³⁰ served as controls. Data are mean \pm s.e.m. of $n=3$ independent experiments; grey symbols show the data mean of technical triplicates of each independent experiment. **e**, Fab CTC binding to the PfCRT 7G8 isoform reduces binding of 370 nM [3 H]CQ, 75 nM [3 H]PPQ or 250 nM [3 H]Arg in a concentration-dependent manner, yielding IC_{50} values of $0.15 \pm 0.02 \mu\text{M}$, $0.20 \pm 0.02 \mu\text{M}$ or $0.18 \pm 0.02 \mu\text{M}$, respectively. Data are mean \pm s.e.m. of $n=3$ independent experiments of technical triplicates. **f**, Time course of 93 nM [3 H]CQ, 75 nM [3 H]PPQ, or 125 nM [3 H]Arg uptake measured with PfCRT 7G8-containing proteoliposomes. Data are mean \pm s.e.m. of $n=2$ independent experiments performed as technical triplicates. Inset, transport kinetics of PfCRT 7G8 for [3 H]CQ and [3 H]Arg in proteoliposomes revealed Michaelis–Menten constants (K_m) of $0.77 \pm 0.15 \mu\text{M}$ and $1.30 \pm 0.20 \mu\text{M}$ with a maximum velocity of transport (V_{\max}) of $92.2 \pm 4.9 \text{ nmol mg}^{-1} \text{ min}^{-1}$ and $189.2 \pm 9.1 \text{ nmol mg}^{-1} \text{ min}^{-1}$ for the uptake of CQ and Arg, respectively. Data show means of $n=2$ independent experiments of technical replicates ($n=4$) and the kinetic constants represent the mean \pm s.e.m. of the global fits. **g**, Uptake of 370 nM [3 H]CQ or 250 nM [3 H]Arg was measured for 1 min in the presence or

absence of 1 μM Fab CTC, 10 μM verapamil, CQ, PPQ or amodiaquine, or 1 mM Leu. Values were normalized to the signal in the absence of the non-radiolabelled compound. Data are mean \pm s.e.m. of $n=3$ independent experiments; grey symbols show the data mean of technical triplicates of each independent experiment. **a–e**, Data are normalized to the specific signal (total counts per minute minus counts per minute in the presence of imidazole) in the absence of the respective non-radiolabelled compound. **h**, Saturation binding of [3 H]CQ (red) or [3 H]PPQ (blue) by indicated PfCRT variants reconstituted in nanodiscs was performed using equilibrium dialysis at pH 5.5. Dissociation constants (K_d) are shown as mean \pm s.e.m. calculated from global nonlinear regression fitting ($n=3$ independent experiments). **i**, Uptake of 100 nM [3 H]CQ (red) or [3 H]PPQ (blue) was measured for 1 min in proteoliposomes containing the indicated PfCRT variants or in control liposomes. Data are means of $n=2$ independent experiments; grey symbols show the data mean of technical replicates ($n=4$) of each independent experiment. **j**, Kinetic characterization of [3 H]CQ (red) or [3 H]PPQ (blue) uptake in proteoliposomes containing the indicated PfCRT variants. The initial rate of transport was measured for periods of 3 s using [3 H]CQ or [3 H]PPQ concentrations ranging from 0.05 to 10 μM . Data (means of $n=2$ independent experiments of technical replicates ($n=4$)) were fitted to the Michaelis–Menten equation. The K_m and V_{\max} values are shown as mean \pm s.e.m. of the fit in Extended Data Table 2b. Transport kinetics of PfCRT 7G8, PfCRT Dd2 or wild-type PfCRT for PPQ or wild-type PfCRT for CQ were not determined due to low signal-to-noise ratios. **k**, [3 H]CQ and [3 H]PPQ CARs at $t=60$ min for edited and reference parasite lines. Data are mean \pm s.e.m. of $n=4–6$ independent experiments performed in duplicate ($n=2$). * $P < 0.05$; ** $P < 0.01$ (using two-tailed Mann–Whitney U -tests). **l**, Percentage parasite survival of *pfCRT*-edited and reference lines following exposure to PPQ for 48 h. Survival was calculated for each line by dividing the parasitaemia of the PPQ-treated parasites with that of the no-drug control. WT, wild type. Data are mean \pm s.e.m. for $n=2–5$ independent assays performed in duplicate ($n=2$). Statistical significance was determined by two-tailed Mann–Whitney U -tests for comparisons in which $n > 2$.

the C350R mutation, which is found in 7G8 parasites and which has previously been proposed to have arisen from local PPQ drug pressure⁹.

First, we expressed the recombinant PfCRT variants 7G8, 7G8 + F145I, 7G8 + C350R and Dd2, as well as the canonical CQ-S and PPQ-S wild-type isoform (from 3D7 parasites). These purified proteins were reconstituted in nanodiscs and tested in equilibrium dialysis experiments for [³H]CQ and [³H]PPQ binding. Results showed a comparable dissociation constant (K_d) across all variants for both drugs at pH 7.5 (Extended Data Table 2b). However, at pH 5.5, which reflects the pH of the acidic digestive vacuole²⁵, we observed an approximately twofold increase in the PPQ⁴⁺ K_d (that is, an approximately twofold lower binding affinity) for the 7G8 + F145I and 7G8 + C350R variants compared with the 7G8 parental isoform (Fig. 3h and Extended Data Table 2b). Testing PPQ uptake in PfCRT proteoliposomes revealed substantially increased PPQ transport rates by the 7G8 + F145I and 7G8 + C350R variants (Fig. 3i, j and Extended Data Table 2b). Minimal PPQ transport was observed with the PPQ-S isoforms 7G8, Dd2 and wild-type PfCRT. Whereas the low transport activity of the 7G8 and Dd2 parental isoforms correlated with the higher PPQ⁴⁺ binding affinity at pH 5.5 (Fig. 3h), this relationship was not observed for wild-type PfCRT (which exhibits a low PPQ⁴⁺ binding affinity and minimal PPQ transport activity). For CQ, results showed robust transport activity with the CQ-R 7G8 and Dd2 parental isoforms, reduced uptake by the 7G8 + F145I and 7G8 + C350R variants, and minimal uptake by wild-type PfCRT (Fig. 3i). For the 7G8 + C350R variant, the higher CQ²⁺ binding affinity (Fig. 3h) paralleled reduced CQ transport (Fig. 3i). However, for the wild-type protein the low CQ transport activity could not be attributed to tight CQ²⁺ binding that would potentially impede the effective dissociation of the drug from the transporter. Instead we propose that the presence of positive K76 in wild-type PfCRT inhibits the required drug–PfCRT interactions within the cavity to enable transport. In summary, these biochemical data provide evidence that both the F145I and C350R mutations can mediate PPQ efflux when introduced into the 7G8 isoform, concomitant with reduced CQ transport. These data also suggest that drug-specific transport, differentially mediated by PfCRT mutations that occur at distinct sites in the central cavity, is the major determinant of resistance.

To correlate these findings to drug susceptibility in the parasite, we genetically edited *pfCRT* in the *P. falciparum* 7G8 line to express the 7G8 + F145I or 7G8 + C350R variants and compared these to control isogenic 7G8 parasites that express the 7G8 isoform. Their characterization revealed a significant increase in the intracellular accumulation ratio (CAR) for CQ after 1 h in 7G8 + F145I and 7G8 + C350R clones, compared to the CQ-R 7G8 and Dd2 lines that showed a minimal CAR (Fig. 3k and Extended Data Table 3). Wild-type (3D7) CQ-S parasites showed high levels of CQ accumulation. For PPQ, the F145I and C350R variants showed a decrease in the CAR compared to the 7G8 parent, although the difference was smaller than for CQ. PPQ accumulation levels were uniformly high for the PPQ-S 7G8, Dd2 and 3D7 parasites. These CAR data are consistent with 72-h parasite growth-inhibition assays that showed reversal of CQ resistance upon introduction of the F145I and C350R mutations into 7G8 parasites, mirrored by increased PPQ IC₅₀ and IC₉₀ values in these variants (Extended Data Table 3). The 7G8 + F145I mutant showed a pronounced biphasic dose–response curve, with IC₉₀ values showing a 188-fold increase compared to the isogenic 7G8 control line. A similar fold increase in PPQ IC₉₀ values was observed in parallel studies that compared the Dd2 + F145I mutant with isogenic Dd2 parasites (Extended Data Fig. 9d). These findings provide evidence that the F145I mutation affords comparable high-grade PPQ resistance at elevated PPQ concentrations in these two genetically and geographically distinct backgrounds⁶ (Extended Data Fig. 9d).

As stated above, PPQ resistance in parasites is evident only at high nanomolar PPQ concentrations^{6,28}. Accordingly, PPQ resistance is best examined using PPQ survival assays, in which synchronized ring-stage parasites (0–6 h after invasion of erythrocytes) are exposed to high concentrations of PPQ for 48 h and the percentage survival of parasites is

measured 24 h later as a percentage of surviving drug-treated parasites compared to mock-treated parasites²⁹. Results from this assay revealed a survival of 9–10% at the standard assay concentration of 200 nM PPQ, and similar levels of survival were maintained at PPQ concentrations of up to 1,600 nM in 7G8 + F145I or 7G8 + C350R parasites (Fig. 3l and Extended Data Table 3). PPQ-S parasites showed 1% survival or less at concentrations of 200 nM or higher. These results are in line with our biochemical studies, which showed increased PPQ transport, accompanied by diminished CQ transport, after the introduction of the Asian F145I and South American C350R mutations into the 7G8 isoform. These findings provide compelling evidence that these PfCRT mutations could affect PPQ efficacy in South America.

To link our structural and functional findings, we calculated the electrostatic potential surfaces of the central cavity for the PfCRT isoforms described above. 7G8 and Dd2 exhibit the most electronegativity in the central cavity, consistent with their highest affinity for PPQ at pH 5.5 (Fig. 3h and Extended Data Fig. 10a). Assuming PPQ⁴⁺ formation at pH 5.5, these variants would feature the strongest electrostatic interaction of PPQ⁴⁺ in the PfCRT cavity. Introducing an arginine into this cavity (7G8 + C350R) caused a reduction in its electronegativity (Extended Data Fig. 10b), which could explain the reduced binding affinity for PPQ and the resulting transport-positive phenotype of this variant (Fig. 3h–j). In the case of PfCRT 7G8 + F145I, however, the electrostatic charge in the central cavity appeared to be unchanged when calculated on the 7G8 model. Molecular dynamics simulations suggested that the F145I mutation induced a conformational change in PfCRT, which involved substantial displacement of TM1, with 145I (TM3) showing greater atomic distance from L83 (TM1) compared to F145 in 7G8 (Extended Data Fig. 10c). TM7 was also notably displaced towards the 145I variant, as shown by the distance to V263. By contrast, marginal movement was observed with the C350R mutation (Extended Data Fig. 10d). These simulations suggest that changes in the cavity shape and size, which result from the F145I mutation, could affect molecular interactions that are essential for the differential transport of CQ and PPQ.

Our analyses highlight a complex mechanism of resistance that is dependent on the gain of CQ²⁺ or PPQ⁴⁺ efflux through specific mutations in the central cavity of PfCRT, which evolved under regional selective drug pressures that began decades ago with the former first-line drug CQ and, more recently, involves PPQ. These mutation-specific changes can differentially influence the electrostatic and other physicochemical properties of the cavity, affecting drug binding and/or the structural dynamics that are associated with transport³⁰. The apparent dichotomy observed between CQ and PPQ for the F145I and C350R mutants, which evolved in CQ-R isoforms and caused CQ resensitization along with a gain of PPQ resistance, highlights the value of extending this research to rapidly emerging PfCRT mutations^{2,7,8}, as a means to identify antimalarial drug combinations that could exert opposing selective pressures on this transporter.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1795-x>.

1. Su, X. Z., Lane, K. D., Xia, L., Sá, J. M. & Wellems, T. E. *Plasmodium* genomics and genetics: new insights into malaria pathogenesis, drug resistance, epidemiology, and evolution. *Clin. Microbiol. Rev.* **32**, e00019 (2019).
2. van der Pluijm, R. W. et al. Determinants of dihydroartemisinin–piperaquine treatment failure in *Plasmodium falciparum* malaria in Cambodia, Thailand, and Vietnam: a prospective clinical, pharmacological, and genetic study. *Lancet Infect. Dis.* **19**, 952–961 (2019).
3. Fidock, D. A. et al. Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6**, 861–871 (2000).

4. Martin, R. E. et al. Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* **325**, 1680–1682 (2009).
5. Blasco, B., Leroy, D. & Fidock, D. A. Antimalarial drug resistance: linking *Plasmodium falciparum* parasite biology to the clinic. *Nat. Med.* **23**, 917–928 (2017).
6. Ross, L. S. et al. Emerging Southeast Asian PfCRT mutations confer *Plasmodium falciparum* resistance to the first-line antimalarial piperaquine. *Nat. Commun.* **9**, 3314 (2018).
7. Hamilton, W. L. et al. Evolution and expansion of multidrug-resistant malaria in Southeast Asia: a genomic epidemiology study. *Lancet Infect. Dis.* **19**, 943–951 (2019).
8. Dhingra, S. K., Small-Saunders, J. L., Ménard, D. & Fidock, D. A. *Plasmodium falciparum* resistance to piperaquine driven by PfCRT. *Lancet Infect. Dis.* **19**, 1168–1169 (2019).
9. Pelleau, S. et al. Adaptive evolution of malaria parasites in French Guiana: reversal of chloroquine resistance by acquisition of a mutation in *pfprt*. *Proc. Natl Acad. Sci. USA* **112**, 11672–11677 (2015).
10. World Health Organization. *World Malaria Report 2018*. <https://www.who.int/malaria/publications/world-malaria-report-2018/en> (2018).
11. Sullivan, D. J. Jr. Quinolines block every step of malaria heme crystal growth. *Proc. Natl Acad. Sci. USA* **114**, 7483–7485 (2017).
12. Dhingra, S. K. et al. A variant PfCRT isoform can contribute to *Plasmodium falciparum* resistance to the first-line partner drug piperaquine. *mBio* **8**, e00303-17 (2017).
13. Lakshmanan, V. et al. A critical role for PfCRT K76T in *Plasmodium falciparum* verapamil-reversible chloroquine resistance. *EMBO J.* **24**, 2294–2305 (2005).
14. Sanchez, C. P. et al. Differences in *trans*-stimulated chloroquine efflux kinetics are linked to PfCRT in *Plasmodium falciparum*. *Mol. Microbiol.* **64**, 407–420 (2007).
15. Paguio, M. F., Cabrera, M. & Roepe, P. D. Chloroquine transport in *Plasmodium falciparum*. 2. Analysis of PfCRT-mediated drug transport using proteoliposomes and a fluorescent chloroquine probe. *Biochemistry* **48**, 9482–9491 (2009).
16. Sanchez, C. P. et al. Phosphomimetic substitution at Ser-33 of the chloroquine resistance transporter PfCRT reconstitutes drug responses in *Plasmodium falciparum*. *J. Biol. Chem.* **294**, 12766–12778 (2019).
17. Renaud, J. P. et al. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Rev. Drug Discov.* **17**, 471–492 (2018).
18. Dominik, P. K. et al. Conformational chaperones for structural studies of membrane proteins using antibody phage display with nanodiscs. *Structure* **24**, 300–309 (2016).
19. Callaghan, P. S., Hassett, M. R. & Roepe, P. D. Functional comparison of 45 naturally occurring isoforms of the *Plasmodium falciparum* chloroquine resistance transporter (PfCRT). *Biochemistry* **54**, 5083–5094 (2015).
20. Quick, M. & Javitch, J. A. Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proc. Natl Acad. Sci. USA* **104**, 3603–3608 (2007).
21. Lekostaj, J. K., Natarajan, J. K., Paguio, M. F., Wolf, C. & Roepe, P. D. Photoaffinity labeling of the *Plasmodium falciparum* chloroquine resistance transporter with a novel perfluorophenylazido chloroquine. *Biochemistry* **47**, 10394–10406 (2008).
22. Bellanca, S. et al. Multiple drugs compete for transport via the *Plasmodium falciparum* chloroquine resistance transporter at distinct but interdependent sites. *J. Biol. Chem.* **289**, 36336–36351 (2014).
23. Sá, J. M. et al. Geographic patterns of *Plasmodium falciparum* drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proc. Natl Acad. Sci. USA* **106**, 18883–18889 (2009).
24. Juge, N. et al. *Plasmodium falciparum* chloroquine resistance transporter is a H⁺-coupled polyspecific nutrient and drug exporter. *Proc. Natl Acad. Sci. USA* **112**, 3356–3361 (2015).
25. Kuhn, Y., Rohrbach, P. & Lanzer, M. Quantitative pH measurements in *Plasmodium falciparum*-infected erythrocytes using pHluorin. *Cell. Microbiol.* **9**, 1004–1013 (2007).
26. Ashcroft, F., Gadsby, D. & Miller, C. Introduction. The blurred boundary between channels and transporters. *Phil. Trans. R. Soc. Lond. B* **364**, 145–147 (2009).
27. Agrawal, S. et al. Association of a novel mutation in the *Plasmodium falciparum* chloroquine resistance transporter with decreased piperaquine sensitivity. *J. Infect. Dis.* **216**, 468–476 (2017).
28. Bopp, S. et al. *Plasmepsin II–III* copy number accounts for bimodal piperaquine resistance among Cambodian *Plasmodium falciparum*. *Nat. Commun.* **9**, 1769 (2018).
29. Duru, V. et al. *Plasmodium falciparum* dihydroartemisinin–piperaquine failures in Cambodia are associated with mutant K13 parasites presenting high survival rates in novel piperaquine in vitro assays: retrospective and prospective investigations. *BMC Med.* **13**, 305 (2015).
30. Shi, L., Quick, M., Zhao, Y., Weinstein, H. & Javitch, J. A. The mechanism of a neurotransmitter:sodium symporter—inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol. Cell* **30**, 667–677 (2008).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Small-scale expression screen of CRT variants

Initial expression studies were performed with twelve CRT variants, including the *P. falciparum* crt sequences 7G8 (AF233064), 3D7 (Q9N623; codon-optimized for *Xenopus laevis*), Dd2 (D5L5S2; codon-optimized for *X. laevis*), *Plasmodium berghei* CRT (PbCRT; Q9GSD8), *Plasmodium knowlesi* CRT (PkCRT; Q9GSD7), *Plasmodium vivax* CRT (PvCRT; Q9GSD3, codon-optimized for *Spodoptera frugiperda*), *Theileria parva* CRT (TpCRT; Q4N5R6), *Theileria annulata* CRT (TaCRT; Q4UDS9) and *Cryptosporidium parvum* CRT (CpCRT; Q7YZ23). To express these in mammalian cells, the sequences were cloned into pFML2³¹ as a GFP fusion with a decahistidine affinity tag at either the 5' or 3' end of the gene. In brief, 1 µg of each construct was diluted into 100 µl opti-MEM (Thermo Fisher Scientific) and mixed with 100 µl opti-MEM containing 4 µg of polyethylenimine (PEI) maximum molecular mass of 40,000 Da (Polysciences). This mixture was added to 1.0×10^6 HEK293T cells (Invitrogen) in a total volume of 2 ml in a 6-well plate (Corning). The cell transfection mixture was incubated at 37 °C for 72 h in 5% CO₂. Transfected cells were then collected, centrifuged at 800g for 10 min at 4 °C and washed once on ice in 1× PBS. Each pellet was resuspended and solubilized in 20 mM HEPES pH 7.5, 200 mM NaCl, 20 mM MgSO₄, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), cOmplete EDTA-free protease inhibitor cocktail (Roche), 10 µg ml⁻¹ DNase I (Roche) and 8 µg ml⁻¹ RNase (Sigma-Aldrich) supplemented with 1% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM) and 0.1% (w/v) cholesterol hemisuccinate (CHS) at 4 °C for 2 h. Insoluble material was removed by ultracentrifugation in a single-angle rotor at 4 °C for 45 min. The supernatant of each sample was subjected to western blot analysis using a rabbit anti-GFP antibody (diluted 1:5,000; Invitrogen) and a secondary goat anti-rabbit IgG conjugated to horseradish peroxidase (HRP) (diluted 1:7,000; Invitrogen). Labelling was detected using a western blot Luminol reagent (Santa Cruz).

Expression and stability were further evaluated by fluorescence-coupled size-exclusion chromatography (FSEC)³². In these assays, each GFP-tagged construct was transfected into HEK293 Freestyle cells (Invitrogen) in the presence of PEI. Transfected cells were incubated at 37 °C for 72 h with 8% CO₂ and 70% humidity. Cell supernatants were prepared for FSEC analysis as described above. DDM-solubilized supernatants were loaded onto a TSKgel G4000SWxl column (Tosoh Bioscience) attached to a Prominence UFLC (Shimadzu) coupled to a RF-10AXL fluorescence detector (Shimadzu). These studies identified the PfCRT 7G8 construct as having the best yield and mono-dispersity.

PfCRT 7G8 protein expression, purification and reconstitution into nanodiscs

The *pfcr*t 7G8 full-length open-reading frame was cloned into the pEG BacMam vector³³ using the Gibson Assembly method³⁴. *pfcr*t was fused at its 3' end with a tobacco etch virus protease cleave site (ENLYFQSYV) and a decahistidine affinity tag followed by a streptavidin affinity tag (WSHPQFEK). The resulting plasmid was transformed into *Escherichia coli* DH10Bac competent cells to generate the bacmid using a bac-to-bac (Invitrogen) protocol. Recombinant P1 baculovirus was transfected into Sf9 cells (Expression System) in the presence of PEI and cultured in ESF 921 protein-free insect cell culture medium (Expression Systems). For protein expression, 100 ml P4 virus was used to infect 1-l cultures of HEK293S GnTi⁻ cells (Invitrogen) at 3×10^6 cells per ml in Freestyle 293 medium (GIBCO) supplemented with 2% FBS (GIBCO). After infection, the cells were incubated at 37 °C for 8 to 10 h in the presence of 8% CO₂ and 70% humidity. We then added 10 mM sodium butyrate (Sigma-Aldrich) to enhance BacMam protein expression. Cells were further incubated at 37 °C for 48 h before collection. Cell pellets were homogenized in low-salt buffer (10 mM HEPES pH 7.5, 10 mM KCl, 10 mM MgCl₂, 0.5 mM PMSF, cOmplete EDTA-free protease inhibitor cocktail, 10 µg ml⁻¹ DNase I, and 8 µg ml⁻¹ RNase) in a glass homogenizer.

Membrane fractions were isolated by ultracentrifugation at 40,000g in a type 45 Ti rotor (Beckman Coulter). Membrane fractions were further homogenized and washed twice with high-salt buffer (10 mM HEPES pH 7.5, 10 mM KCl, 10 mM MgCl₂, 1 M NaCl, 0.5 mM PMSF, cOmplete protease inhibitor cocktail, 10 µg ml⁻¹ DNase I and 8 µg ml⁻¹ RNase) in a glass homogenizer followed by ultracentrifugation. The washed membrane fractions were resuspended again by homogenizing in buffer comprising 20 mM HEPES pH 7.5, 200 mM NaCl and 0.5 mM PMSF and cOmplete protease inhibitor cocktail, and stored at -80 °C until use.

The thawed membrane fraction was solubilized by adding DDM with CHS in a 10:1 (w/w) ratio to a final concentration of 1% (w/v) detergent, as previously described³⁵, and incubated at 4 °C for 2 h with gentle agitation. Insoluble material was removed by ultracentrifugation at 40,000g in a type 45 Ti rotor (Beckman Coulter) at 4 °C for 30 min. The supernatant was filtered (0.22-µm filters, Millipore), placed in a Falcon tube containing pre-equilibrated Ni²⁺-NTA resin (Qiagen) in the presence of 20 mM imidazole, and incubated at 4 °C overnight with gentle rotation. The resin was washed with 10 column volumes of buffer comprising 20 mM HEPES pH 7.5, 200 mM NaCl, 60 mM imidazole, 0.1% DDM and 0.01% CHS. Bound protein was eluted with buffer consisting of 20 mM HEPES pH 7.5, 200 mM NaCl, 200 mM imidazole, 0.05% DDM and 0.005% CHS. The eluted protein was incorporated into lipid nanodiscs with a 1:400:10 molar ratio of protein:1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol):membrane-scaffold protein 1D1 (MSP1D1). This mixture was incubated at 4 °C for 2.5 h with gentle agitation. Reconstitution was initiated by removing detergent by incubating with Bio-beads (Bio-Rad) at 4 °C overnight with constant rotation. Bio-beads were removed and the nanodisc reconstitution mixture was bound again to Ni²⁺-NTA resin for at 4 °C for 2 h to remove free nanodiscs. The resin was washed with 10 column volumes of wash buffer (20 mM HEPES pH 7.5, 200 mM NaCl and 20 mM imidazole) followed by 4 column volumes of elution buffer (20 mM HEPES pH 7.5, 200 mM NaCl and 200 mM imidazole). The eluted protein was further purified by loading on a Superdex 200 Increase 10/300 GL size-exclusion column (GE Healthcare Life Sciences) in gel-filtration buffer (20 mM HEPES pH 7.0 and 150 mM NaCl).

Identification of PfCRT-specific Fabs using phage display

For the Fab screen, PfCRT 7G8 was reconstituted into nanodiscs using biotinylated MSP1D1. Before nanodisc assembly, MSP1D1 was chemically biotinylated as previously described^{18,36,37}. Efficient biotinylation was confirmed using a pull-down assay with streptavidin-coated paramagnetic particles (Promega). Biopanning was performed using the Fab Library E^{38,39} that expresses a subset of Fabs on the phage surface. Binding was assayed in selection buffer (25 mM HEPES pH 7.4, 150 mM NaCl and 1% bovine serum albumin). In the first round, biopanning was performed manually using 200 nM of PfCRT 7G8-MSP1D1 nanodiscs immobilized onto magnetic beads. Following three washes with selection buffer, the resuspended beads enriched for phage-expressing PfCRT-specific Fabs were used to infect log-phase *E. coli* XL-1 Blue bacteria. Phages were amplified overnight in 2× YT medium supplemented with ampicillin (100 µg ml⁻¹) and M13-KO7 helper phage (10⁹ PFU ml⁻¹). To increase the stringency of selection, four additional rounds of biopanning were performed with decreasing target concentrations of 150 nM, 75 nM, 40 nM and 20 nM. For each round, the amplified phage pool from each preceding round was used as the input. For rounds 2–5, biopanning was performed semi-automatically using a Kingfisher magnetic beads handler (Thermo Fisher Scientific). To reduce the presence of non-specific binders, the phage pools for rounds 2–5 were precleared with 100 µl of streptavidin particles. For all rounds we used empty, non-biotinylated MSP1D1 nanodiscs as soluble competitors in excess, at a constant concentration of 1.5 µM. In rounds 2–5, bound phage particles were removed by elution from magnetic beads following a 15-min incubation step with 1% Fos-choline-12 prepared in selection buffer.

Single-point phage ELISA to validate Fab binding to PfCRT

The initial validation for selection was performed by single-point phage ELISA using individual clones from rounds 4 and 5. All ELISA experiments were performed in 96-well plates (Nunc) coated with $2 \mu\text{g ml}^{-1}$ neutravidin and blocked with selection buffer. Colonies of *E. coli* XL1-Blue bacteria containing phagemids were used to inoculate $400 \mu\text{l}$ $2\times$ YT medium supplemented with $100 \mu\text{g ml}^{-1}$ ampicillin and 10^9 PFU ml^{-1} M13-KO7 helper phages. Phages were amplified overnight in 96-well deep-well blocks at 37°C with shaking at 280 rpm. Amplified phages were diluted tenfold into selection buffer and assayed against either PfCRT-loaded or empty biotinylated nanodiscs. Target proteins were immobilized at room temperature for 30 min followed by incubation with phage dilutions for 30 min. Bound phage particles were detected with TMB substrate (Thermo Fisher Scientific) following a 30-min incubation with HRP-conjugated anti-M13 monoclonal antibody (GE Healthcare). The reaction was quenched with 1.0 M HCl and the absorbance was measured at 450 nm. Wells containing empty nanodiscs were used to determine non-specific binding.

Fab expression and purification

Specific binders from phage ELISA were selected based on their signal/background ratio¹⁸ and were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing facility. Unique clones were sub-cloned in pRH2.2 using the In-Fusion Cloning kit (Takara Bio). Sequence-verified Fab expression vectors were then transformed into *E. coli* BL21-Gold competent cells (Agilent), and grown overnight in $2\times$ YT broth supplemented with $100 \mu\text{g ml}^{-1}$ ampicillin at 37°C . On the next day, cultures were inoculated into fresh YT broth, grown at 37°C until optical density at 600 nm reached 0.8, and then induced for 4 h at 37°C . Cells were collected by centrifugation and stored at -80°C until use. Cell pellets were resuspended in buffer (20 mM HEPES pH 7.4, 150 mM NaCl and 0.5 mM MgCl_2) supplemented with 1 mM PMSF and $1 \mu\text{g ml}^{-1}$ DNase I (Gold Biosciences). These preparations were sonicated (Branson Sonifier) until complete lysis was achieved. Lysates were incubated at 60°C for 30 min to eliminate potential Fab proteolysed fragments, subsequently cooled on ice and cleared by centrifugation. Supernatants were filtered and then loaded onto a 5-ml HiTrap MabSelect SuRe column (GE Healthcare) equilibrated with wash buffer (20 mM HEPES pH 7.4 and 500 mM NaCl). The column was washed 10 times with wash buffer and Fabs were eluted with 0.1 M acetic acid. Fractions containing Fabs were loaded onto a 1-ml Resource S column (GE Healthcare) equilibrated with buffer A (50 mM sodium acetate, pH 5.0). Following washing with 10 column volumes of buffer A, Fabs were eluted by a linear 0–50% gradient with buffer B (50 mM sodium acetate, pH 5.0 and 2 M NaCl). Fractions containing Fabs were evaluated for purity by SDS–PAGE and subsequently pooled and dialysed overnight in 20 mM HEPES pH 7.4 and 150 mM NaCl.

Assessment of Fab binding affinity for PfCRT

To estimate apparent binding affinity, we performed multi-point ELISA with purified Fabs¹⁸, which were diluted serially threefold starting from a $3\text{-}\mu\text{M}$ maximum concentration. ELISA plates were coated and blocked as described above for the phage ELISA studies. Then, 50 nM of biotinylated PfCRT 7G8-MSP1D1 nanodiscs was immobilized on the plates by a 30-min incubation, followed by washing. Fab dilutions were added to wells containing immobilized PfCRT and allowed to bind for 30 min. Bound Fab was detected with TMB substrate following a 30-min incubation with HRP-conjugated mouse anti-human IgG F(ab')₂ monoclonal antibody (Jackson). Reactions were quenched with 1.0 M HCl and the absorbance was measured at 450 nm. A_{450} values were plotted against the \log_{10} of the Fab concentration. EC_{50} values were then calculated in GraphPad Prism 8 using a variable slope model and assuming a sigmoidal dose response.

PfCRT 7G8 complex formation with the Fab CTC

Size-exclusion chromatography (SEC) fractions of purified PfCRT in nanodiscs were incubated with the Fab CTC on ice for 2 h in a 1:3 molar ratio of protein to Fab. The PfCRT–Fab complex was concentrated and filtered, and then loaded on a Superdex 200 Increase 10/300 GL size-exclusion column in gel-filtration buffer (20 mM HEPES pH 7.0 and 150 mM NaCl).

Negative-stain electron microscopy

To determine the sample quality and success of the Fab binding, purified protein was diluted to 0.01 mg ml^{-1} (for PfCRT without Fab) or 0.005 mg ml^{-1} (for PfCRT with Fab) and applied onto copper grids (Ted Pella). These grids were overlaid by a thin (approximately 1.5 nm) layer of continuous carbon that had been plasma-cleaned (Gatan Solarus) for 30 s using a mixture of H_2 and O_2 . Afterwards, filter paper (Whatman 4) was used to remove the protein solution. Then, $3 \mu\text{l}$ of 2% uranyl formate was added and immediately removed by absorbing with filter paper—this was repeated seven times. The grid was imaged on either a Tecnai T12 microscope (FEI) (for PfCRT without Fab) or a Tecnai TF20 microscope (FEI) (for PfCRT with Fab). Both microscopes were equipped with a Tietz F416 CCD camera (Tietz) with 1.23 \AA or 1.10 \AA per pixel, respectively, using the Leginon software package⁴⁰. In total, 166 and 87 images were collected for PfCRT without and with Fab, respectively, and all images were processed using the Appion software package⁴¹ to obtain two-dimensional classes with Relion 2.1^{42,43}. The micrographs showed good particle dispersion. Two-dimensional class averages showed that the Fab addition resulted in a clear fiducial for particle alignment (Extended Data Fig. 2).

Single-particle cryo-EM vitrification and data acquisition

Purified PfCRT–Fab complex was concentrated to 1.56 mg ml^{-1} using a 30-kDa concentrator (Amicon). In brief, $2.5 \mu\text{l}$ of sample was added to a plasma-cleaned (Gatan Solarus) 1.2/1.3- μm holey gold grid (Quantifoil UltrAuFoil) and blotted using filter paper on one side for 2 s using the Leica GP plunger system, before plunging immediately into liquid ethane for vitrification. The plunger was operating at 6°C with $>80\%$ humidity to minimize evaporation and sample degradation. Images were recorded on a Titan Krios electron microscope (FEI) equipped with a C_s corrector and K2 summit direct detector (Gatan) operating at 0.5175 \AA per pixel (calibrated using an AAV2 dataset) in counting mode using the Leginon software package⁴⁰. An energy filter slit width of 20 eV was used during the collection and was aligned automatically every hour using Leginon. Data collection was performed using a dose of around $91.56 \text{ e}^- \text{ \AA}^{-2}$ across 80 frames (75 ms per frame) at a dose rate of approximately 4.0 e^- per pixel per s, using a set defocus range of $-1.2 \mu\text{m}$ to $-1.8 \mu\text{m}$. A $100\text{-}\mu\text{m}$ objective aperture was used. In total, 3,377 micrographs were recorded over a single 2.5-day collection using an image beam shift data-collection strategy⁴⁴. Ice thickness was monitored after every fourth exposure using the Leginon zero-loss peak algorithm⁴⁵ and was determined to have a mean \pm s.d. of $23.1 \pm 9.1 \text{ nm}$.

Data processing

Video frames were aligned using MotionCor2⁴⁶ with 3 by 3 patches, a grouping of 3 and a B-factor of 100 using the Appion software package⁴¹. Micrograph contrast transfer function (CTF) estimations were performed using both CTFFind4⁴⁷ and GCTF⁴⁸, and the best estimates based on confidence were selected using the Appion software package. Micrographs that had an estimated resolution at 0.5 confidence level of worse than 10 \AA were removed, resulting in 3,297 micrographs. DoG picker⁴⁹ was used to pick 682,520 particles (extracted binned by 2), which were transferred into Relion 2.1^{42,43} for two-dimensional classification. Two-dimensional class averages that showed clear structural details of PfCRT were used as templates for template-based picking using Gautomatch (<https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch>). To avoid

missing any particles, we chose a lenient threshold that resulted in 1,095,304 initial picks. These were pared down to 183,241 particles after Relion two-dimensional classification. Two rounds of CryoSPARC v.2⁵⁰ ab initio (using two models) were then performed to further refine the particle stack to 35,682 particles, which yielded a resolution of 8.27 Å after three-dimensional homogeneous refinement. Particles were then re-centred and re-extracted using Relion and per-particle CTF estimation using GCTF was then performed. The particle stack was brought back into CryoSPARC for one round of ab initio refinement (using two models). The best set of particles and model was then used for non-uniform refinement. This resulted in a 3.7 Å map from 17,034 particles.

The `mag_distortion_correct` software⁵¹ was then used on the micrographs to correct for magnification anisotropy of 1.3%, which was determined previously using gold-replica grating using the `mag_distortion_estimate` software⁵¹. The resulting particles were then subjected to three rounds of CryoSPARC non-uniform refinement, re-centering and re-extracting in Relion, with removal of overlapped particles centred no more than 20 pixels from each other, and CTF refinement using `cisTEM`⁵². This resulted in a 3.6 Å density map from 17,030 particles.

To optimize the alignment of frames, video frames were re-aligned using `MotionCor2` with 2 by 2 patches, a grouping of 3, correction for in-frame motion, and a B-factor of 150 for global frame alignment and 50 for local frame alignment. Visual inspection of the resulting micrographs showed that certain micrographs fared better using the new frame alignment parameters, while some fared worse than with the 3 by 3 patches described above. To determine the best frame-alignment parameter on a per-particle basis, particles were extracted from micrographs using both conditions. These particles were then subjected to a CryoSPARC non-uniform refinement and one round of `cisTEM` refinement. Thereafter, we chose the alignment parameter that produced the highest score. This resulted in 7,191 particles coming from micrographs aligned with 2 by 2 patches, and 9,714 particles coming from micrographs aligned with 3 by 3 patches. Particle-polishing algorithms that require neighbouring particles to aid in alignment were tested but gave consistently worse results, probably because of the low average number of particles per micrograph (around 5.3) caused by the high magnification used and sizeable number of free Fab contaminants.

Next, signal subtraction using Relion was performed. Subtraction of just the nanodisc did not improve the overall resolution. However, subtraction of both the nanodisc and constant domain of the Fab improved the resolution of the reconstruction, after CryoSPARC non-uniform refinement, to resolutions of 3.3 Å overall and 3.2 Å for PfCRT alone. This final map was sharpened using `phenix.auto_sharpen`⁵³, which automatically selected `b_iso` sharpening to `high_resolution` cutoff as the algorithm to use. Overall `b_sharpen` applied was 28.3 Å², final `b_iso` obtained was 47.3 Å² and the high-resolution cut-off was 3.2 Å. All conversions between Relion, CryoSPARC and `cisTEM` were performed using the `pyem` script (D. Asarnow, unpublished materials, <https://github.com/asarnow/pyem>).

Structural model building and refinement

To build the PfCRT model, we first used the Chimera⁵⁴ program to approximately segment out the PfCRT part of the map from the nanodisc and Fab density. De novo model building for the PfCRT primary sequence in the cryo-EM map was initiated using Rosetta⁵⁵. After two rounds of Rosetta model building, the partial model was brought into Coot⁵⁶ for manual model building. For Fab model building, the Fab portion of the deposited yeast nucleoprotein complex (PDB 4XMM)⁵⁷ was used as a starting template, as it had the same scaffold backbone. An extra electron density between PfCRT helices 1, 9 and 10 was not modelled by protein density and was likely of lipid origin. Out of the lipids present in the nanodisc, CHS produced the best density fit and was therefore incorporated into the model.

Subsequently, model adjustment and refinement were performed iteratively in Coot and Phenix, with the statistics examined using

Molprobit⁵⁸ until no further improvements were observed. The final map and model were then validated using: (1) EMRinger⁵⁹ to compare map to model; (2) ResMap⁶⁰ to calculate map local resolution; and (3) a program suite with three-dimensional Fourier shell correlations (FSCs)⁶¹. To calculate the FSC values, we used a cut-off criterion⁶² of 0.143. The three-dimensional FSC program suite⁶¹ was used to calculate the degree of directional resolution anisotropy. Map-to-model FSCs were also calculated by first converting the model to a map using the Chimera `molmap` function at Nyquist resolution (2.07 Å). A mask was generated from this map using Relion (after low-pass filtering to 8 Å, extending by 1 pixel and applying a cosine-edge of 3 pixels) and was then applied to the density map. Map-to-model FSCs were calculated using EMAN⁶³.

Model analysis

A cavity search using the Solvent Extractor from Voss Volume Voxelator server⁶⁴ was performed using an outer-probe radius of 10 Å and inner-probe radius of 3 Å. We then used a Dali server⁶⁵ to search for other PDB structures with a similar fold. This search produced top results for Vrg4 (PDB 5OGE and 5OGK)⁶⁶, YddG (PDB 5I20)⁶⁷, TPT (PDB 5Y78 and 5Y79)⁶⁸, *Zea mays* CST (PDB 6I1R)⁶⁹ and *Mus musculus* CST (PDB 6OH4)⁷⁰, with Z scores ranging from 26.7 to 16.0. Thereafter, there was a large drop in Z score for the next best hit at 8.5 (the chloride pumping rhodopsin, PDB 5B2N) that visually did not superimpose well on PfCRT (the root mean square deviation of superimposition was 8.1 Å).

Amino acid sequence alignments

For conservation analysis in apicomplexan parasites, we obtained 11 orthologues of CRT from OrthoMCL-DB⁷¹. Their amino acid sequences, in addition to the PfCRT Dd2 and wild-type isoforms, were then aligned using MUSCLE⁷², resulting in 13 orthologues in total. For the sequences used for Extended Data Fig. 7, the sequences used are *P. falciparum* PfCRT strain 7G8 (UNIPROT W7FI62) and its orthologues PfCRT Dd2 (UNIPROT F5CEB4), wild-type PfCRT (the canonical 3D7 wild-type sequence; UNIPROT Q9N623), *P. reichenowi* CRT (PrCRT; UNIPROT AOA2P9D9K2), *P. vivax* CRT strain Sal-1 (PvCRT; UNIPROT Q9GSD3), *P. knowlesi* CRT strain H (PkCRT; UNIPROT Q9GSD7), *P. berghei* CRT strain ANKA (PbCRT; UNIPROT Q9GSD8), *P. chabaudi* CRT strain chabaudi (PcCRT; UNIPROT Q7Z0V9), *Babesia microti* CRT strain RI (BmCRT; UNIPROT AOA1N6LY67), *T. annulata* CRT strain Ankara (TaCRT, UNIPROT Q4UDS9); *Eimeria tenella* CRT (EtCRT; UNIPROT U6L1M8), *Toxoplasma gondii* CRT (TgCRT; UNIPROT S8EU26) and *Cryptosporidium hominis* CRT strain TU502 (ChCRT; UNIPROT AOA0S4THJ3). Multiple sequence alignments were displayed using ESPrpt⁷³.

Computational studies

All calculations were performed using the Schrödinger molecular modelling suite (version 2019-1). The PfCRT protein structure solved herein was analysed using the Protein Preparation Wizard⁷⁴. In this step, force field atom types and bond orders were assigned, missing atoms including hydrogens were added, tautomer/ionization states were assigned, Asn, Gln and His residues were flipped to optimize the hydrogen bond network, and a constrained energy minimization was performed. For mutation modelling, including to generate models of Dd2 and WT (3D7), point mutations were introduced into the PfCRT 7G8 structure using the Residue and Loop Mutation tool. The refinement step consisted of a local minimization in implicit solvent with Prime⁷⁵.

For simulations of molecular dynamics, we used the same protocol to simulate the 7G8 isoform and the different single amino acid variants. First, the protein was aligned so that the membrane plane corresponded to the *xy* plane of the coordinate system, using the OPM (Orientation of Proteins in Membranes) database⁷⁶. Then, the protein was embedded in a 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) membrane model, included in an orthorhombic box of simple point-charge (SPC) water solvent with a 10 Å buffer on each

dimension, and neutralized by adding the appropriate number of Cl⁻ ions, using the System Builder panel. The simulations of molecular dynamics were performed using Desmond⁷⁷. Each simulation started with a system relaxation with a default protocol that was followed by a 250-ns production run in an NPgT ensemble, at a temperature of 300 K and a pressure of 1.01325 bar. The resulting trajectories were analysed using the Simulation Event Analysis tool to monitor distances of interest and the frames for each trajectory were clustered using the Desmond Trajectory tool.

The Poisson–Boltzmann electrostatic potential surface panel, which uses the adaptive Poisson–Boltzmann solver method, was used to calculate the electrostatic potential surfaces. The solute and solvent dielectric constants were set to 1 and 80, respectively, solvent radius was 1.4 Å, temperature was 298 K and grid extension was 5 Å. The open-to-cytosol homology model of PfCRT was created using Maestro's Multiple Sequence Alignment tool. Leveraging the inverted structural repeats of PfCRT that are related by two-fold pseudo-symmetry, the sequence of each TM helix was aligned and mapped onto the structure of the related TM helix, as previously carried out for other transporters with inverted structural repeats^{67,78–80}. The positions of the loops and JM helices in the open-to-cytosol conformation could not be defined owing to a lack of homology in these regions. The individually aligned TM helices were merged to give the final model, which was further refined using restrained minimization from the Protein Preparation Wizard panel⁷⁴.

Production of *pfcr*t gene-edited parasites

7G8 parasites were engineered to express *pfcr*t variants that added the F145I or C350R mutations, or the 7G8 reference sequence, into the 7G8 coding sequence, using a previously published method based on zinc-finger nuclease-mediated gene editing^{9,81}. Edited clones were recovered by limiting dilution. Parasite lines were authenticated by sequencing the *pfcr*t alleles and whole-genome sequencing of the 7G8 parental line (provided by T. Wellems, NIAID/NIH). Testing for mycoplasma revealed no contamination.

[³H]CQ and [³H]PPQ accumulation assays in cultured parasites

This protocol has been previously described in detail⁸². In brief, sorbitol-synchronized parasite cultures of recombinant 7G8, 7G8 + F145I, 7G8 + C350R, Dd2 and 3D7 (expressing wild-type *pfcr*t) were collected as early trophozoites (24–26 h after invasion) and magnet-purified (VarioMACS, Miltenyi Biotec). These lines are referred to herein as 7G8, 7G8 + F145I, 7G8 + C350R, Dd2 and wild-type (3D7). The pure trophozoite-infected erythrocytes were eluted at 37 °C in fresh 1× PBS containing 2 mM EDTA and supplemented with 0.5% bovine serum albumin. The eluate was then centrifuged (1,500 rpm, 5 min), and the pellet was washed and resuspended in bicarbonate-free RPMI 1640 medium (supplemented with 25 mM HEPES, 10 mM glucose and 0.2 mM hypoxanthine and adjusted to pH 7.4) at 37 °C. The exact incubation cell numbers were determined by counting an aliquot of the resuspended cells using a haemocytometer. Uptake measurements began with the addition of 250 µl of trophozoite suspension, at an estimated haematocrit of 30,000–120,000 cells per µl, to an equal volume of 10 nM [³H]CQ (60 Ci mmol⁻¹; American Radiolabelled Chemicals) or 40 nM [³H]PPQ (15 Ci mmol⁻¹; American Radiolabelled Chemicals) in bicarbonate-free medium to yield final [³H]CQ and [³H]PPQ concentrations of 5 nM and 20 nM, respectively. Parallel control measurements were conducted with uninfected erythrocytes at the same haematocrit. After 1 h of incubation at 37 °C in a water bath, duplicate 200 µl aliquots of the suspensions were transferred into 1.5-ml Eppendorf tubes containing 300 µl of dibutyl phthalate (Sigma Aldrich, 1.04 g ml⁻¹) and centrifuged immediately (13,000 rpm; 2 min) to sediment the cells through the oil, thereby terminating [³H]-drug uptake. From each duplicate, 100 µl of supernatant was transferred into a vial containing 5 ml of scintillation fluid (Cytoscint, MP Biomedicals) to determine the external drug

concentration [Drug]_{external}. The pellets were digested by overnight incubation at 55 °C with 100 µl tissue solubilizer (NCS-II, GE Healthcare) diluted 1:2 in absolute ethanol, then bleached the next day with 30% hydrogen peroxide (25 µl) followed by addition of 25 µl glacial acetic acid to block luminescence. Digested pellets were transferred into a vial containing 10 ml of scintillation cocktail and measured in a scintillation counter to determine the internal drug concentration [Drug]_{internal}. After adjusting for non-specific uptake by subtracting measurements from the uninfected erythrocyte controls, the CAR was calculated as the ratio of [Drug]_{internal}/[Drug]_{external}, normalized to 1 × 10⁶ infected erythrocytes and assuming a mean volume of a trophozoite-infected erythrocyte of 75 femtolitres⁸³.

Parasite PPQ survival assays

PPQ survival assays were carried out as previously described with slight modifications⁸⁴. In brief, double sorbitol-synchronized ring-stage parasites (0–6 h after invasion) were seeded at 1% parasitaemia and 1% haematocrit in 96-well flat-bottom plates containing a twofold, eight-point PPQ drug dilutions starting at 1,600 nM. Parasites were incubated with the drug for 48 h at 37 °C. The drug was then removed by washing the plates four times with complete medium using a TECAN Freedom EVO 100 liquid handler. The assay plates were further incubated for 24 h before parasitaemia was measured using an Accuri C6 flow cytometer by double staining with SYBR Green I and MitoTracker Deep Red (to label DNA and the mitochondria, respectively). Percentage survival was calculated for each line by dividing the parasitaemia of the PPQ-treated parasites by the parasitaemia of the no-drug control. Statistical significance was determined by nonparametric, two-tailed Mann–Whitney *U*-tests using GraphPad Prism 8 software.

Parasite drug-susceptibility assays

Parasite susceptibility to antimalarial drugs was measured by incubating asynchronous asexual blood-stage parasites to a range of drug concentrations (twofold, ten-point) starting at 0.3% parasitaemia and 1% haematocrit in 96-well plates. The assay plates were incubated at 37 °C for 72 h in an environment with 5% O₂/5% CO₂/90% N₂. Final parasitaemias were measured using flow cytometry. IC₅₀ values were calculated by nonlinear regression analysis. Drug-susceptibility assays were performed with CQ, monodesethyl-chloroquine (the in vivo metabolite of CQ) and PPQ. Statistical significance was determined by nonparametric, two-tailed Mann–Whitney *U*-tests using GraphPad Prism 8 software.

Binding studies

All radiolabelled compounds were purchased from American Radiolabelled Chemicals. Unless other concentrations are indicated, binding experiments used 125 nM [³H]Arg (40 Ci mmol⁻¹), 370 nM [³H]CQ (5.45 Ci mmol⁻¹) or 75 nM [³H]PPQ (15 Ci mmol⁻¹). Scintillation proximity-based assay (SPA) binding²⁰ of 200 ng of PfCRT 7G8, reconstituted into nanodiscs, was assayed in buffer composed of 50 mM Tris/MES, pH 7.5, 5% glycerol, 0.1 mM TCEP, 0.1% PEI in the presence of 125 µg of YSi-His tag SPA beads (Perkin Elmer), and in the presence or absence of indicated reagents or Fab. Then, 800 mM imidazole, which competes with the His tag (fused to PfCRT) for binding to the copper-coated YSi SPA beads, was used to determine the non-proximity background signal. Concentrations of radiolabelled compounds were chosen in pilot experiments to obtain a signal-to-noise ratio ≥ 5, and specific binding was determined by subtracting the counts per minute in the presence of imidazole from the counts per minute measured in its absence. Control binding experiments of [³H]CQ were carried out using nanodiscs containing 100 ng of the unrelated *Aquifex aeolicus* bacterial amino acid transporter LeuT. As the SPA technology prevented reliable binding experiments at pH < 7.5, binding of [³H]CQ (1 Ci mmol⁻¹) or [³H]PPQ (2.5 Ci mmol⁻¹) was performed with equilibrium dialysis using the HTD96b dialysis 96-well apparatus and 12,000–14,000 MWCO membranes in 50 mM Tris/MES, pH 7.5 or pH 5.5, 5% glycerol, 0.1 mM TCEP for

4 h at 4 °C with 4.2 pmol of the indicated PfCRT variant in nanodiscs⁸⁵. Nonlinear regression analysis of the data was performed with GraphPad Prism 7. Kinetic constants represent the mean ± s.e.m. of the fit.

Transport measurements

Purified PfCRT variants were reconstituted in preformed liposomes made of *E. coli* total lipids:CHS 94:6 (w/w) at a protein-to-lipid ratio of 1:150 (w/w). The lumen of the proteoliposomes was composed of 100 mM K_P, pH 7.5 and 2 mM β-mercaptoethanol. Uptake of [³H]Arg (40 Ci mmol⁻¹), [³H]CQ (5.45 Ci mmol⁻¹) or [³H]PPQ (15 Ci mmol⁻¹) was performed by diluting PfCRT-containing proteoliposomes (40 ng PfCRT per assay) in 50 μl of 100 mM Tris/MES, pH 5.5 in the presence or absence of the indicated compounds. Unless indicated otherwise, 1 μM valinomycin was added to the reaction to generate a K⁺ diffusion potential-driven membrane potential. The reactions were stopped after the indicated time periods by the addition of ice-cold 100 mM K_P, pH 6.0, 100 mM LiCl and filtered through 0.45-μm nitrocellulose filters (Millipore). Filters were dried and incubated in scintillation cocktail, and the radioactivity captured on the filters was counted in a Hidex SL300 scintillation counter. Specific radioactivity of all compounds was confirmed by counting known amounts of each radiolabel by the determination of the decay per minute to transform the decay per minute of the samples into pmol. Unspecific binding of all compounds with the nitrocellulose filters was determined by measuring the mock uptake in the absence of liposomes or proteoliposomes, and these values (determined for each experiment) were used to calculate the uptake activity in liposomes or proteoliposomes. The specific uptake was determined by subtracting the time-dependent accumulation of the tested compounds in control liposomes (lacking PfCRT) from that measured in PfCRT-containing proteoliposomes. Kinetic constants (K_m and V_{max}) were determined by measuring the initial rates of [³H]Arg, [³H]CQ or [³H]PPQ for periods of 3 s by fitting the data to the Michaelis–Menten equation in GraphPad Prism 7 and represent the mean ± s.e.m. of the fit.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All raw video frames, micrographs, the particle stack and relevant metadata files have been deposited into EMPIAR, with accession code EMPIAR-10330. The electron density map has been deposited into EMD, with accession code EMD-20806. The model has been deposited in the PDB, with accession code 6UKJ. All data are available in the paper or Supplementary Information.

31. Assur, Z., Hendrickson, W. A. & Mancia, F. Tools for coproducing multiple proteins in mammalian cells. *Methods Mol. Biol.* **801**, 173–187 (2012).
32. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
33. Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585 (2014).
34. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
35. Wright, D. J., O'Reilly, M. & Tisi, D. Engineering and purification of a thermostable, high-yield, variant of PfCRT, the *Plasmodium falciparum* chloroquine resistance transporter. *Protein Expr. Purif.* **141**, 7–18 (2018).
36. Paduch, M. et al. Generating conformation-specific synthetic antibodies to trap proteins in selected functional states. *Methods* **60**, 3–14 (2013).
37. Dominik, P. K. & Kosiakoff, A. A. Phage display selections for affinity reagents to membrane proteins in nanodiscs. *Methods Enzymol.* **557**, 219–245 (2015).
38. Fellouse, F. A., Wiesmann, C. & Sidhu, S. S. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc. Natl Acad. Sci. USA* **101**, 12467–12472 (2004).
39. Fellouse, F. A. et al. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J. Mol. Biol.* **373**, 924–940 (2007).
40. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
41. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
42. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
43. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
44. Cheng, A. et al. High resolution single particle cryo-electron microscopy using beam-image shift. *J. Struct. Biol.* **204**, 270–275 (2018).
45. Rice, W. J. et al. Routine determination of ice thickness for cryo-EM grids. *J. Struct. Biol.* **204**, 38–44 (2018).
46. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
47. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
48. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
49. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
50. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
51. Grant, T. & Grigorieff, N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.* **192**, 204–208 (2015).
52. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
53. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
54. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
55. Wang, R. Y.-R. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).
56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
57. Stuwe, T. et al. Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015).
58. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
59. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
60. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
61. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
62. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
63. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
64. Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38**, W555–W562 (2010).
65. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
66. Parker, J. L. & Newstead, S. Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature* **551**, 521–524 (2017).
67. Tsuchiya, H. et al. Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature* **534**, 417–420 (2016).
68. Lee, Y. et al. Structure of the triose-phosphate/phosphate translocator reveals the basis of substrate specificity. *Nat. Plants* **3**, 825–832 (2017).
69. Nji, E., Gulati, A., Qureshi, A. A., Coincon, M. & Drew, D. Structural basis for the delivery of activated sialic acid into Golgi for sialylation. *Nat. Struct. Mol. Biol.* **26**, 415–423 (2019).
70. Ahuja, S. & Whorton, M. R. Structural basis for mammalian nucleotide sugar transport. *eLife* **8**, e45221 (2019).
71. Chen, F., Mackey, A. J., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* **34**, D363–D368 (2006).
72. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
73. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
74. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimou, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27**, 221–234 (2013).
75. Jacobson, M. P. et al. A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367 (2004).
76. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
77. Bowers, K. J. et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proc. the ACM/IEEE Conference on Supercomputing (SC06)* (2006).
78. Liao, J. et al. Structural insight into the ion-exchange mechanism of the sodium/calcium exchanger. *Science* **335**, 686–690 (2012).
79. Mancuso, R., Gregorio, G. G., Liu, Q. & Wang, D. N. Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. *Nature* **491**, 622–626 (2012).
80. Forrest, L. R. (Pseudo)-symmetrical transport. *Science* **339**, 399–401 (2013).
81. Straimer, J. et al. Site-specific genome editing in *Plasmodium falciparum* using engineered zinc-finger nucleases. *Nat. Methods* **9**, 993–998 (2012).

82. Sanchez, C. P., Stein, W. & Lanzer, M. *Trans* stimulation provides evidence for a drug efflux carrier as the mechanism of chloroquine resistance in *Plasmodium falciparum*. *Biochemistry* **42**, 9383–9394 (2003).
83. Saliba, K. J., Horner, H. A. & Kirk, K. Transport and metabolism of the essential vitamin pantothenic acid in human erythrocytes infected with the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **273**, 10190–10195 (1998).
84. Dhingra, S. K. et al. Global spread of mutant PfCRT and its pleiotropic impact on *Plasmodium falciparum* multidrug resistance and fitness. *mBio* **10**, e02731-18 (2019).
85. Quick, M., Shi, L., Zehnpfennig, B., Weinstein, H. & Javitch, J. A. Experimental conditions can obscure the second high-affinity site in LeuT. *Nat. Struct. Mol. Biol.* **19**, 207–211 (2012).
86. Miller, K. R. et al. T cell receptor-like recognition of tumor in vivo by synthetic antibody fragment. *PLoS ONE* **7**, e43746 (2012).
87. Wu, T. T. & Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **132**, 211–250 (1970).
88. Hohn, M. et al. SPARX, a new environment for cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
89. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
90. Francis, S. E., Sullivan, D. J. Jr & Goldberg, D. E. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu. Rev. Microbiol.* **51**, 97–123 (1997).

Acknowledgements We thank B. Rice, E. Eng, L. Kim, K. Jordan, M. Kopylov, V. Dandey, H. Wei, S. Dallakyan, C. Negro, S. Gabryszewski, S. Mukherjee, B. Riegel, O. Clarke and Y. Chen for their helpful contributions. This work was supported by NIH grants (R01 AI147628 to F.M., D.A.F. and

M.Q.; R35 GM132120 and R01 GM111980 to F.M.; R37 AI50234 and R01 AI124678 to D.A.F.; R01 GM119396 to M.Q.; R01 AI506312 and AI111962 to P.D.R.; R01 GM117372 to A.A.K.; T32 HL120826 to J.K.; P41 GM103310 to C.S.P. and B.C.; the Agency for Science, Technology and Research Singapore (to Y.Z.T.); the Simons Foundation (SF349247 to C.S.P. and B.C.); and NYSTAR (to C.S.P. and B.C.). Some of the work was performed at the Center for Membrane Protein Production and Analysis (COMPPA; P41 GM116799 to Wayne Hendrickson) and at the National Resource for Automated Molecular Microscopy at the Simons Electron Microscopy Center (P41 GM103310), both located at the New York Structural Biology Center.

Author contributions J.K., with help from L.M.H. and S.I.G., performed protein expression and purification. S.K.E., K.N., and A.A.K. identified the Fabs. Y.Z.T. and J.K. produced and analysed the cryo-EM data and built the model with help from B.C. and C.S.P. M.Q. and A.L.W. performed the biochemical assays. K.J.W., S.K.D., J.O., P.D.R. and D.A.F. conducted mutational analyses. K.J.W. and J.V. performed computational studies, and S.K.D. and J.O. performed parasite gene editing and characterization. M.Q., D.A.F. and F.M. designed experiments and wrote the paper with J.K., Y.Z.T., K.J.W. and P.D.R. K.J.W. and S.K.E. contributed equally as second authors.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1795-x>.

Correspondence and requests for materials should be addressed to M.Q., D.A.F. or F.M.

Peer review information *Nature* thanks Shangyu Dang, Leann Tilley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

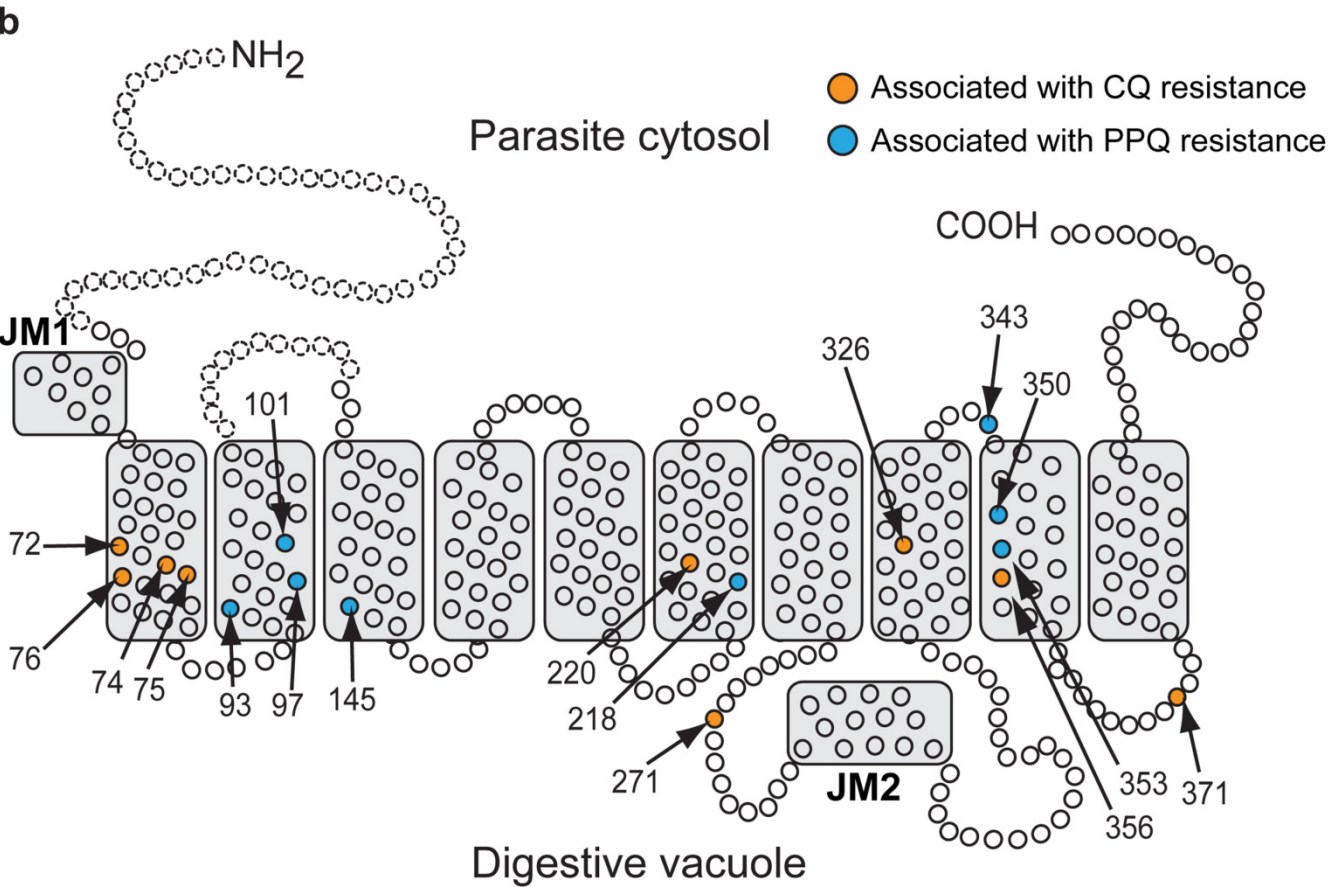
Reprints and permissions information is available at <http://www.nature.com/reprints>.

a

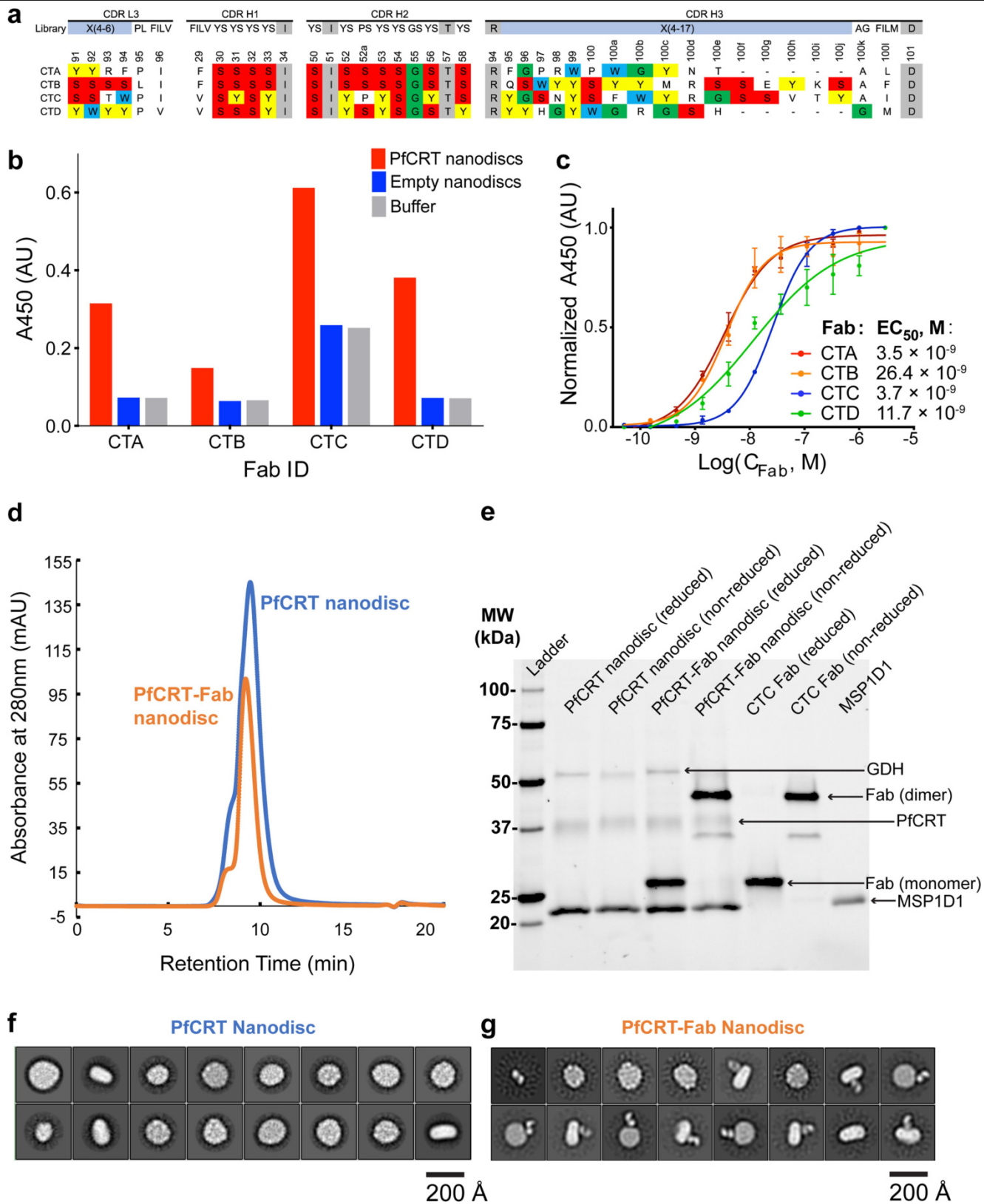
			PfCRT haplotype											Drug phenotype	
Haplotype	Region	Mutations	72	74	75	76	145	220	271	326	350	356	371	CQ	PPQ
3D7 (wild-type)	Africa	0	C	M	N	K	F	A	Q	N	C	I	R	S	S
7G8	S. America, W. Pacific	5	S	--	--	T	--	S	--	D	--	L	--	R	S
7G8+C350R	French Guiana (S. America)	6	S	--	--	T	--	S	--	D	R	L	--	S	R (low)
Dd2	SE Asia	8	--	I	E	T	--	S	E	S	--	T	I	R	S
Dd2+F145I	Cambodia	9	--	I	E	T	I	S	E	S	--	T	I	S ^a	R

Mutations in the reference Dd2 allele are shown in light grey. Other mutations are in dark grey. S, sensitive; R, resistant.

^aSensitized to CQ compared to parental line but not fully sensitive.



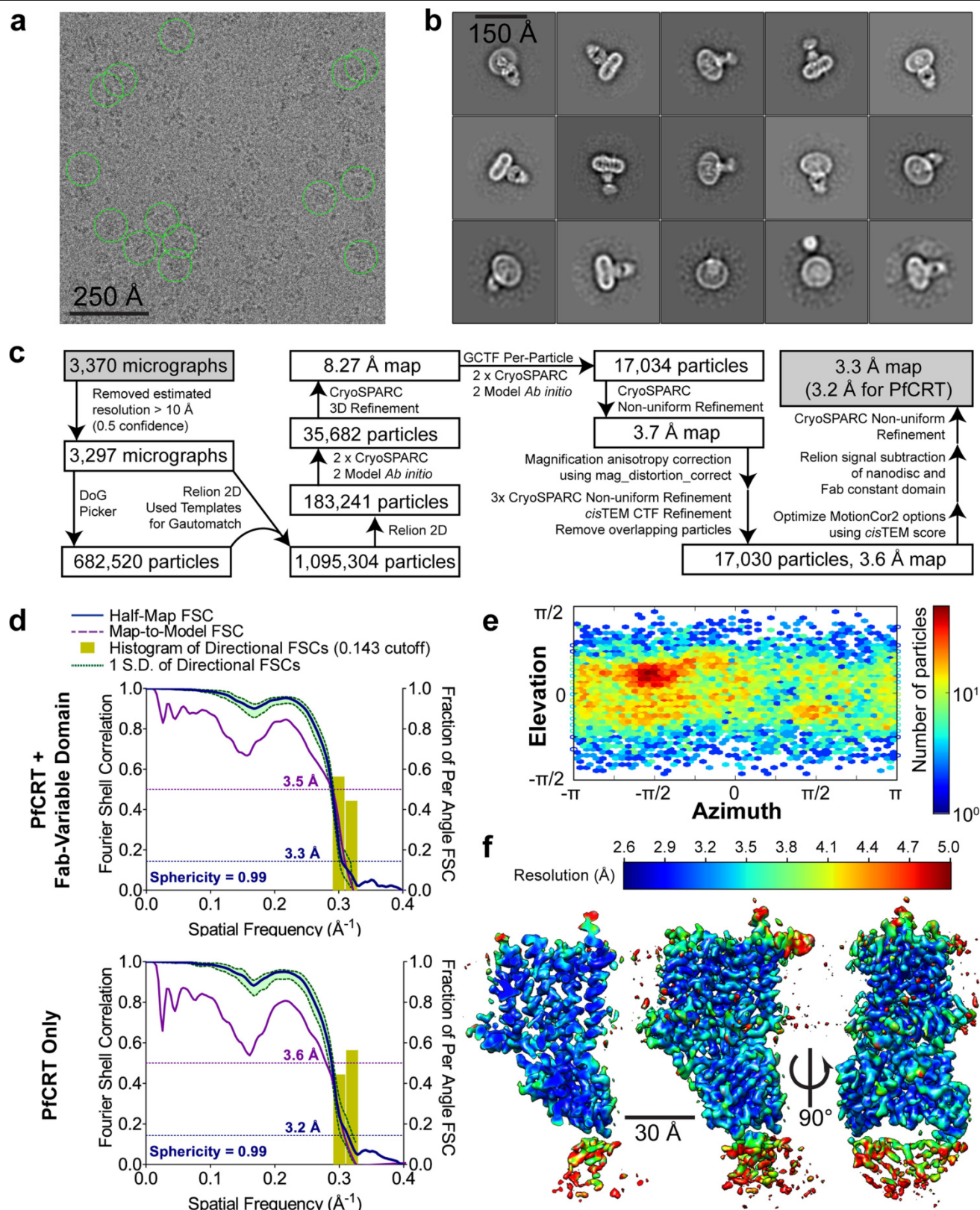
Extended Data Fig. 1 | Major PfCRT haplotypes and location of residues involved in CQ or PPQ drug-resistance phenotypes. **a**, The PfCRT canonical CQ-S, PPQ-S 3D7 haplotype, the CQ-R 7G8 (South America and Western Pacific) and Dd2 (Southeast Asia) haplotypes, and the 7G8 + C350R and Dd2 + F145I variants that have emerged in *P. falciparum* parasites in malaria-endemic areas and are associated with PPQ resistance. **b**, Localization of the mutated residues in PfCRT, based on the solved PfCRT structure. Most mutated residues localize within or near the boundary of one of the 10 TM helices. Extended Data Table 2a provides additional information about the mutated residues in PfCRT observed in the field or obtained in drug-pressured cultured parasites.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Identification of PfCRT-specific Fabs and preparation of purified PfCRT with or without Fabs in nanodiscs. **a**, Complementarity-determining region (CDR) sequences of unique Fabs bioassayed for binding to recombinant PfCRT 7G8 incorporated into MSP1D1 nanodiscs. Fabs were selected following multiple rounds of enrichment from the phage display library E^{38,39,86}. Ser-Tyr-Gly-Trp (SYGW) residues, which play a dominant part in antigen recognition, are highlighted by colour code (red, yellow, green and blue, respectively) and are numbered according to the Kabat system⁸⁷. CTC was selected to form a stable PfCRT 7G8–Fab complex. **b**, Single-point ELISA quantification of the binding of phage-displayed Fab to PfCRT-incorporated biotinylated nanodiscs, empty nanodiscs or buffer (empty wells), measured at 450 nm absorbance ($n = 1$). **c**, EC₅₀ evaluation of purified Fab binding to PfCRT incorporated into biotinylated nanodiscs, showing high-affinity binding for

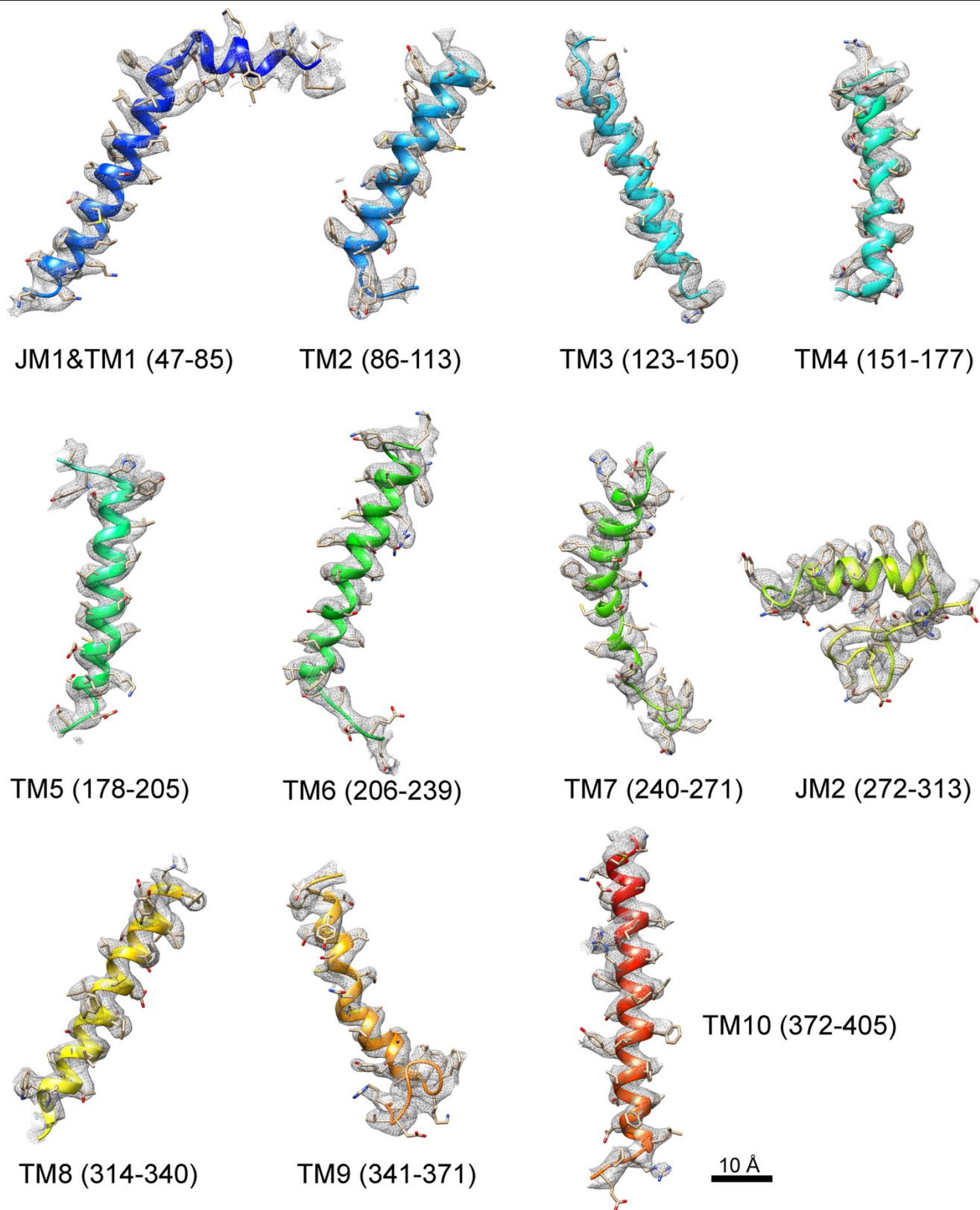
the Fab CTC (3.7 nM). Data are mean \pm s.d. of $n = 3$ independent experiments. **d**, High-performance liquid chromatography profile of PfCRT 7G8 with or without bound Fab CTC. **e**, SDS–PAGE gel of pooled and concentrated SEC fractions from PfCRT 7G8 with or without Fab CTC. The contaminant glutamate dehydrogenase (GDH; present as a left shoulder in **d**) was excluded from single-particle analyses. MSP1D1 is a membrane-scaffold protein used to assemble the nanodiscs. The identity of PfCRT and GDH was confirmed using mass spectrometry. **f**, Representative negative-stain two-dimensional class averages from Relion⁴² two-dimensional classification of nanodisc-incorporated PfCRT without Fab. **g**, Representative negative-stain two-dimensional class averages from Relion two-dimensional classification of nanodisc-incorporated PfCRT with bound Fab CTC.



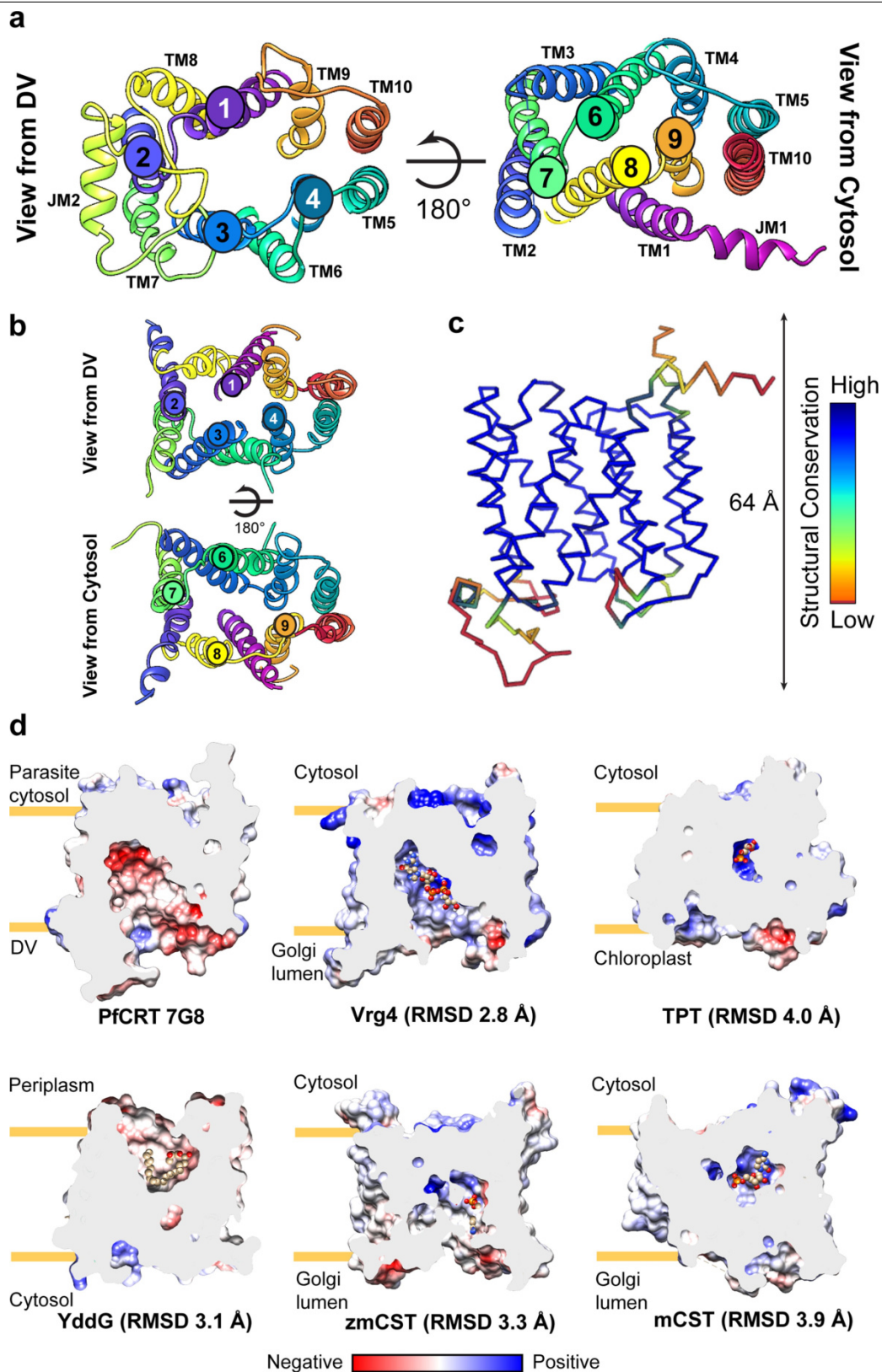
Extended Data Fig. 3 | Cryo-EM analysis of the PfCRT 7G8-Fab CTC complex.

a, Representative micrograph (0.5175 Å per pixel, 1.65 μm defocus). Example picked particles contributing to the final reconstruction are circled in green. **b**, Representative two-dimensional class averages from Relion two-dimensional classification. **c**, Flowchart of cryo-EM data processing and refinement of the PfCRT 7G8-Fab CTC complex. See Methods for details. **d**, FSC curves for PfCRT 7G8 complexed with the Fab CTC variable domain as well as for PfCRT alone. Data show the half-map (blue) and map-to-model (purple) resolutions (at 0.143

and 0.5 cut-offs, respectively), with embedded histograms of directional resolutions sampled evenly over the three-dimensional FSC⁶¹ (yellow). The corresponding sphericity values are indicated. **e**, Euler angle distribution plot of the final three-dimensional reconstruction from CryoSPARC v.2.50. **f**, Local resolution⁸⁸ display of unsharpened reconstructions of PfCRT complexed with the variable domain of Fab CTC, in orthogonal views, sliced through the density for the first view.



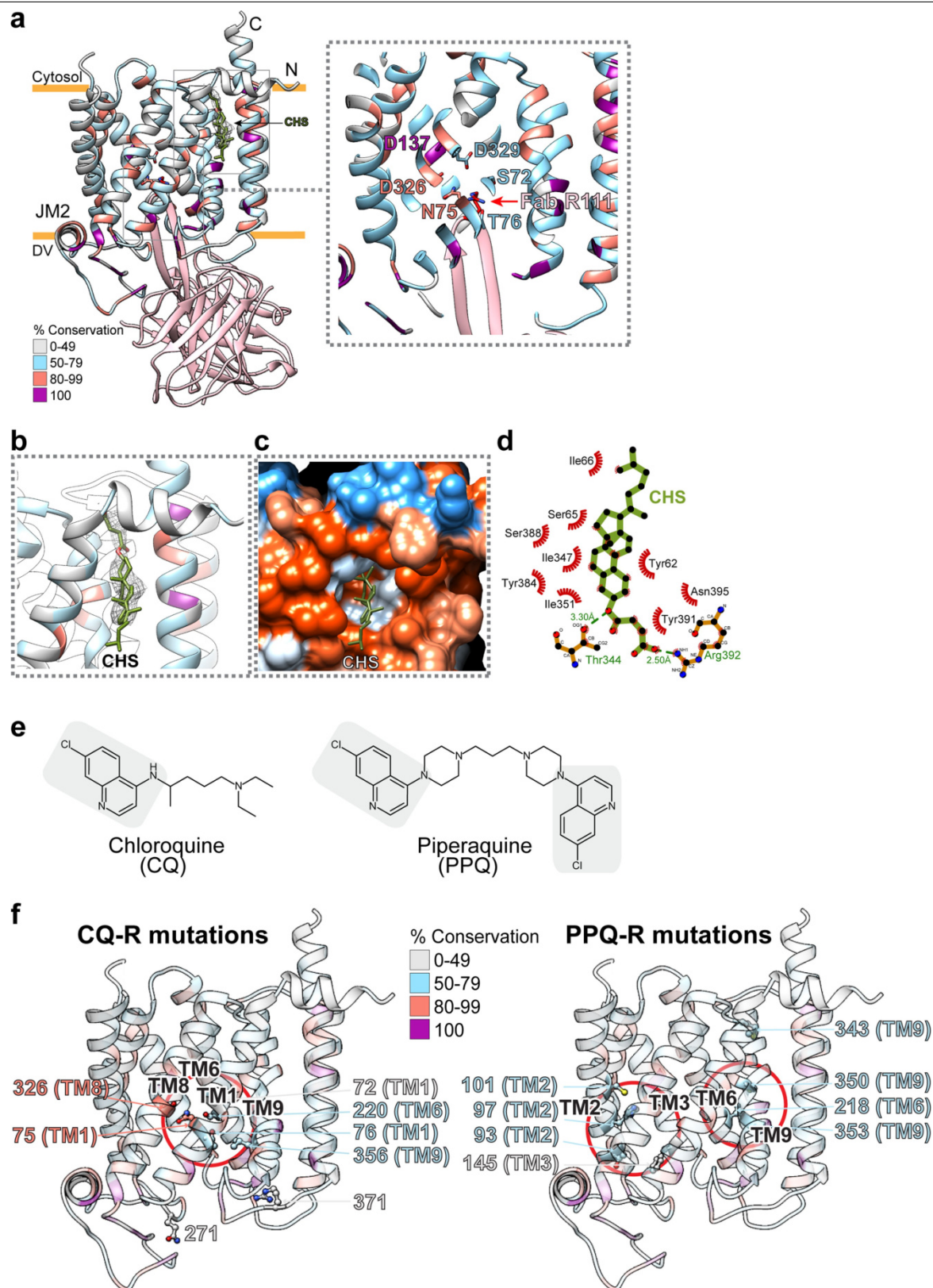
Extended Data Fig. 4 | Fit of cryo-EM density with model. Cryo-EM densities (mesh) are superimposed on TM and JM helices of the PfCRT model. The model is rendered as a cartoon, coloured in rainbow.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | PfCRT symmetrical arrangement, topology of TM helices and structural comparison with other DMT superfamily members. **a**, View of PfCRT from opposite directions of the digestive vacuole and cytosol, with labelling of the four helices closest to the centre (thus contributing to the cavity). **b**, Model of the TM helices in the PfCRT 7G8 structure in the open-to-cytosol conformation. **c**, Structural conservation of PfCRT compared with the DMT family members Vrg4 (a GDP-mannose transporter; PDB 5OGE and 5OGK)⁶⁶, YddG (an amino acid transporter; PDB 5I20)⁶⁷, TPT (a triose-phosphate/phosphate translocator; PDB 5Y78 and 5Y79)⁶⁸, and *Z. mays* CST

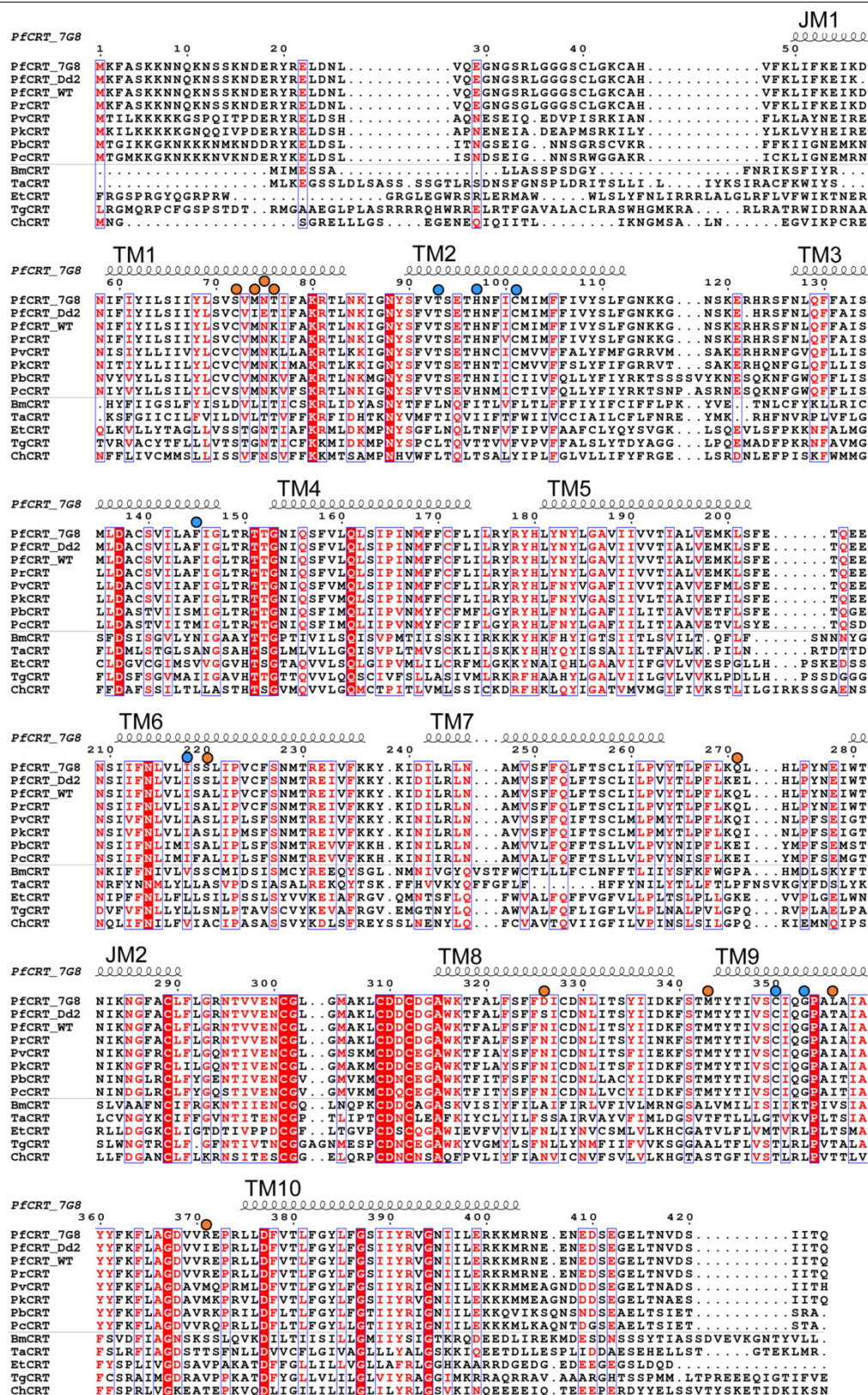
and *M. musculus* CST (two CMP-sialic acid transporters; PDB 6I1R and 6OH4, respectively)^{69,70}. Conservation profiles were generated using the Dali server⁶⁵. **d**, The electrostatic potential surface representation of slices reveals the cavities of these six proteins in their solved structural states. Note the highly negatively charged (red) residues in the PfCRT cavity, contrasting with the positively charged (blue) or neutral residues in the other DMT transporters (shown as slices to emphasize the individual cavities). Ligands for the non-PfCRT transporters are shown as a ball and stick representation.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | The structure of the PfCRT 7G8–Fab CTC complex shows the sequence conservation, localization of CHS, chemical structure of CQ and PPQ, and conservation of CQ-R and PPQ-R mutations. **a**, PfCRT and the bound Fab are rendered in cartoon. PfCRT is coloured according to sequence conservation (see Extended Data Fig. 7) and the Fab is coloured in pink. Inset shows a magnified view of the interaction between PfCRT 7G8 and the Fab CTC. **b**, PfCRT cleft formed by JM1, TM1, TM9 and TM10 demonstrates CHS placement. **c**, PfCRT is rendered as a surface coloured by hydrophobicity, from orange (hydrophobic) to blue (hydrophilic). **d**, Two-dimensional diagram

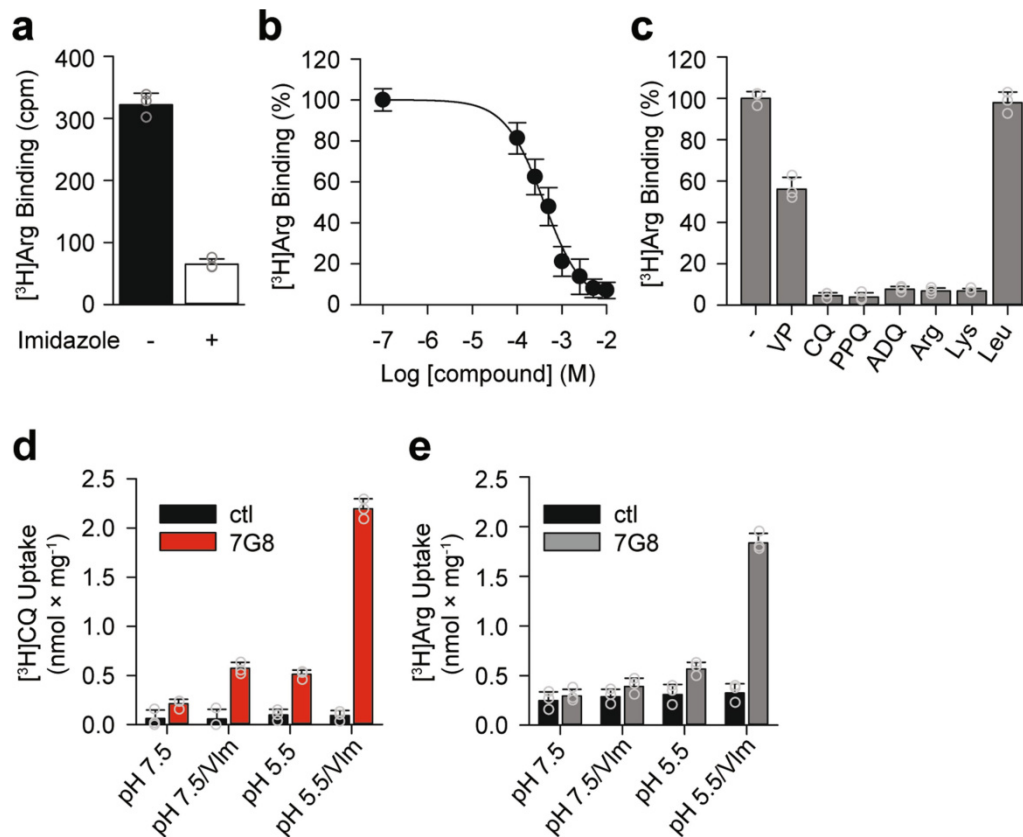
of PfCRT–CHS interactions observed in the structure, generated using LigPlot^{v.2.1}⁸⁹. **e**, Chemical structures of CQ and PPQ, with the 4-aminoquinoline rings shaded. **f**, Location of CQ (left) and PPQ (right) resistance-associated mutations (with side chains rendered as sticks and spheres), modelled onto the PfCRT 7G8 structure. Models are colour-coded according to sequence conservation derived from 11 apicomplexan species (Extended Data Fig. 7). TM helices associated with CQ or PPQ resistance are labelled and the areas highlighted with coloured circles.



Extended Data Fig. 7 | See next page for caption.

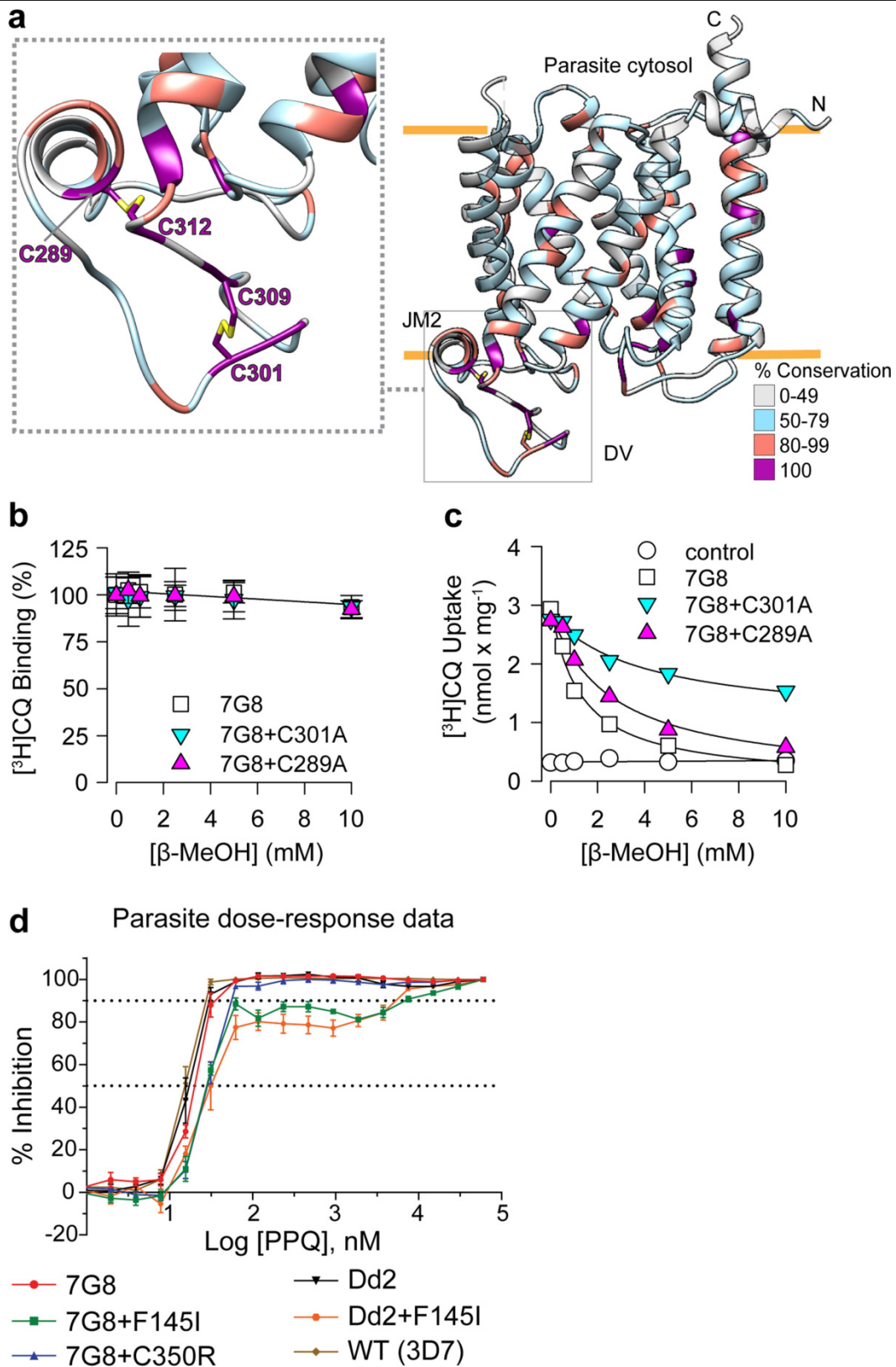
Extended Data Fig. 7 | Sequence alignment and secondary structure of PfCRT. Sequences of different PfCRT isoforms and other CRT homologues were aligned using MUSCLE⁷² and displayed using ESPript⁷³. Sequences of ten orthologues of CRT in other *Plasmodium* or other apicomplexan parasites were obtained from OrthoMCL-DB⁷¹. The sequences used are PfCRT 7G8 (UNIPROT W7FI62) and its orthologues PfCRT Dd2 (UNIPROT F5CEB4) and wild-type PfCRT (the canonical 3D7 wild-type sequence; UNIPROT Q9N623), *P. reichenowi* CRT (PrCRT; UNIPROT A0A2P9D9K2), *P. vivax* CRT strain Sal-1 (PvCRT; UNIPROT Q9GSD3), *P. knowlesi* CRT strain H (PkCRT; UNIPROT Q9GSD7), *P. berghei* CRT strain ANKA (PbCRT; UNIPROT Q9GSD8), *P. chabaudi* CRT strain chabaudi (PcCRT; UNIPROT Q7Z0V9), *B. microti* CRT strain RI (BmCRT; UNIPROT A0A1N6LY67), *T. annulata* CRT strain Ankara (TaCRT, UNIPROT Q4UDS9); *E. tenella* CRT (EtCRT; UNIPROT U6L1M8), *T. gondii* CRT

(TgCRT; UNIPROT S8EU26) and *C. hominis* CRT strain TU502 (ChCRT; UNIPROT A0A0S4THJ3). The secondary structure of PfCRT is shown as a cartoon above the alignments with residue numbering corresponding to the PfCRT 7G8 reference, along with the positioning of the variant residues indicated in Extended Data Fig. 1 (sharing the same colour scheme). The highly conserved cysteine residues that probably form disulphide bonds are C289–C312 and C301–C309. The degree of conservation was calculated by considering all sequences including *P. falciparum* 3D7 (wild type) but excluding the 7G8 and Dd2 variants. Residues conserved (that is, identical or similar) in at least 10 out of 13 species are indicated in red text. Residues conserved in all species are in white text with red highlighting. None of the mutations associated with CQ or PPQ resistance mapped to residues that are fully conserved across these apicomplexan species.



Extended Data Fig. 8 | Binding and transport assays for PfCRT 7G8. **a**, Total binding of 125 nM [³H]Arg to 100 ng of nanodisc-incorporated PfCRT 7G8 was measured in the absence (–) or presence (+) of 800 mM imidazole (which competes with the His-tagged PfCRT 7G8 isoform for binding to the copper-coated YSi SPA beads). Data are mean ± s.e.m. (*n* = 3 independent experiments); grey symbols show the data mean of technical replicates (*n* = 3) of each independent experiment. **b**, Isotopic dilution of 125 nM [³H]Arg with non-radiolabelled Arg revealed a log(*EC*₅₀) value of -3.40 ± 0.033 (corresponding to 397 μM). We note that the primary source of Arg in parasitized red blood cells is from the proteolysis of haemoglobin, which in its native state⁹⁰ as a tetramer is present at approximately 5 mM. Data are mean ± s.e.m. (*n* = 3 independent experiments of technical triplicates). **c**, Binding of 125 nM [³H]Arg in the presence or absence of 10 μM verapamil (VP), 1 μM CQ or PPQ, 0.1 μM

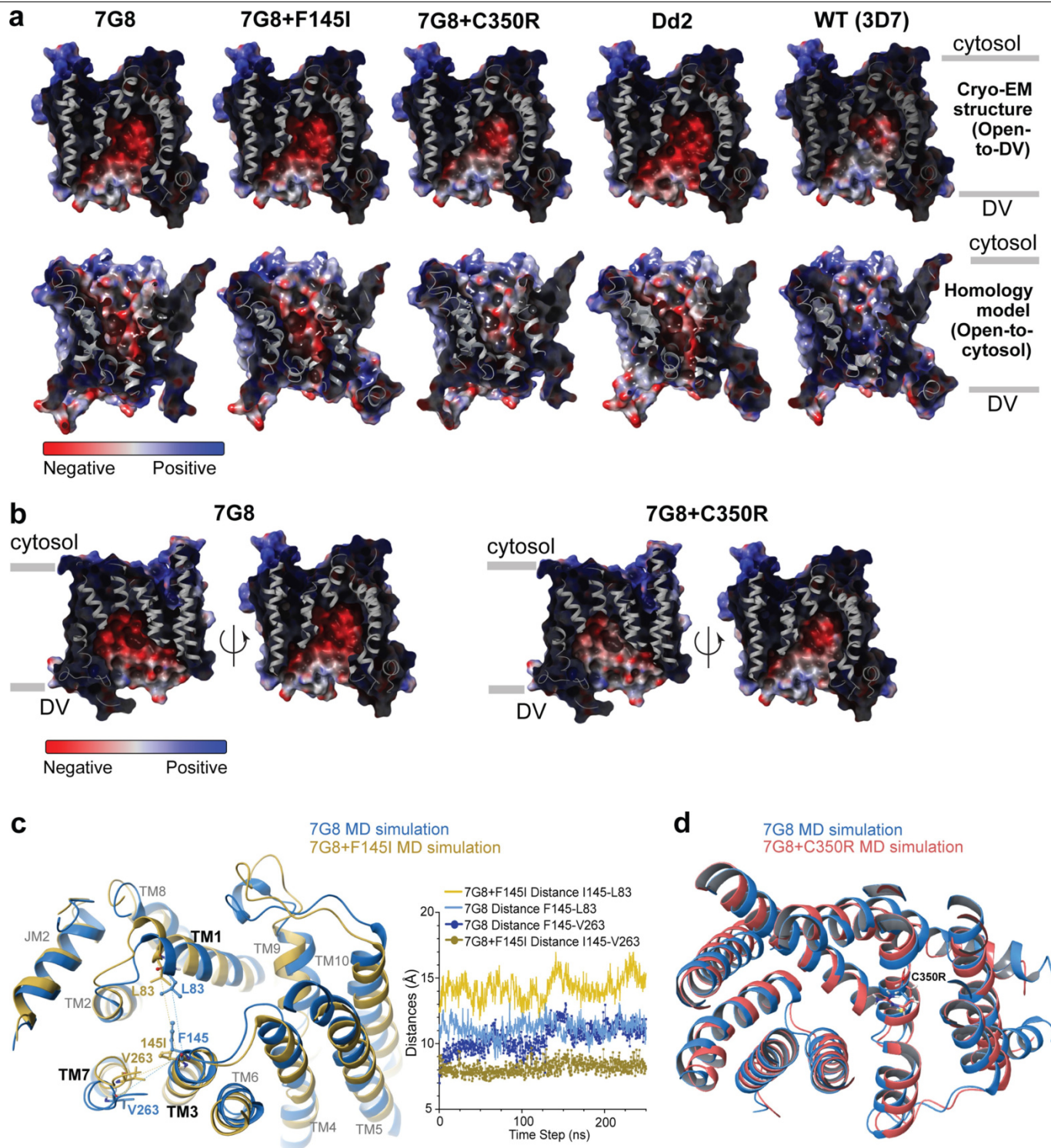
amodiaquine (ADQ), or 1 mM Arg, Lys or Leu. Data were normalized to the signal in the absence of the respective non-radiolabelled compound and are mean ± s.e.m. (*n* = 3 independent experiments); grey symbols show the data mean of technical triplicates of each independent experiment. **d, e**, Uptake of 370 nM [³H]CQ (**d**) or 250 nM [³H]Arg (**e**), was measured for 1-min periods in PfCRT 7G8-containing proteoliposomes preloaded with 100 mM KPi, pH 7.5 diluted in buffer composed of 50 mM Tris/MES, pH 5.5 or pH 7.5 with or without the K⁺ ionophore valinomycin (Vlm; 5 μM). ctl, control. The valinomycin-mediated K⁺ efflux proceeding down its concentration gradient generated an inside-negative membrane potential. Empty liposomes lacking PfCRT 7G8 served as controls. Data are mean ± s.e.m. (*n* = 3 independent experiments); grey symbols show the data mean of technical triplicates of each independent experiment.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Role of conserved cysteine residues in CQ binding and transport, and parasite PPQ dose–response data. **a**, Cysteine residues that probably form disulphide bonds on the loop adjacent to JM2 and connecting to TM8 are shown in yellow and rendered as sticks. Their locations are shown as an inset of the overall structure and are coloured by conservation (as per Extended Data Fig. 7). **b**, Effect of reducing conditions on the binding of [^3H]CQ by PfCRT 7G8, 7G8 + C289A and 7G8 + C301A. SPA binding of 100 nM [^3H]CQ was performed with 200 ng of each protein variant reconstituted into nanodiscs at pH 7.5 in the presence of increasing concentrations of β -mercaptoethanol. Data (shown as mean \pm s.e.m. of $n = 3$ independent experiments with technical triplicates) were normalized with regard to the activity measured for PfCRT 7G8 in the absence of the reducing agent. **c**, Effect of the reducing agent β -mercaptoethanol on [^3H]CQ transport by PfCRT 7G8, 7G8 + C289A or 7G8 + C301A. Uptake of 100 nM [^3H]CQ was measured for 1-min periods in control liposomes and proteoliposomes containing the indicated PfCRT

variants, in the presence of increasing concentrations of β -mercaptoethanol. Data are means of $n = 2$ independent experiments performed as $n = 4$ technical replicates and were subjected to nonlinear regression fitting using hyperbolic decay models with either two (for 7G8 and 7G8 + C289A) or three (including 7G8 + C301A) parameters. Kinetic constants are shown as mean \pm s.e.m. of the fits. The IC_{50} (concentration of β -mercaptoethanol yielding half-maximal reduction of transport) for 7G8 and 7G8 + C289A were 1.24 ± 0.14 mM and 2.58 ± 0.39 mM, respectively. For 7G8 + C301A the IC_{50} was 4.17 ± 0.16 mM with a remaining activity of $60.2 \pm 8.2\%$ (compared to the activity in the absence of the reducing agent). **d**, PPQ dose–response data for *pfcr*t-edited and control parasite lines, generated from 72-h growth inhibition assays. Data are mean \pm s.e.m. from $n = 4$ independent assays performed in duplicate ($n = 2$). Statistical significance was determined using two-tailed Mann–Whitney *U*-tests.



Extended Data Fig. 10 | Electrostatic potential surfaces of isoform-specific PfCRT cavities and simulations of molecular dynamics. **a**, Electrostatic surfaces for the solved open-to-digestive-vacuole conformation for 7G8 and modelled isoforms, predicted at pH 5. Images are presented as a vertical slice through the transporter, showing net charges in the cavity and locations of TM helices. The row below shows the electrostatic surfaces for the homology models of these PfCRT isoforms, illustrated in their open-to-cytosol configuration, predicted at pH 7.0. **b**, Surface representation of the electrostatic potential of the central cavity of 7G8 and the 7G8 + C350R variant,

shown in two different orientations with red and blue indicating negatively and positively charged residues, respectively. **c**, Simulations of molecular dynamics on the 7G8 structure (without the Fab) over 250-ns trajectories, establishing the equilibrium positions of protein side chains and distances between position 145 and residues on proximal helices. **d**, Simulations of molecular dynamics on the 7G8 structure (without the Fab) over 250-ns trajectories, establishing the equilibrium positions of protein side chains between position 350 and residues on proximal helices, and showing marginal movement between the C350 and the 350R isoforms.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

PfcRT with Fab-Variable Domains		
	(EMD-20806)	(PDB 6UKJ)
Data collection and processing		
Magnification	215,000	
Voltage (kV)	300	
Electron exposure (e-/Å ²)	91.56	
Exposure time (s)	6	
Dose rate (e-/pixel/s)	4	
Defocus range (µm)	1.5 ± 0.23	
Pixel size (Å)	0.5175	
Symmetry imposed	C1	
Number of micrographs	3,377	
Initial particle images (no.)	183,241	
Final particle images (no.)	17,030	
Map resolution (Å)	3.3	
FSC threshold	0.143	
Map resolution range (Å)		
Local	2.6-5.6	
Directional	3.1-3.3	
Sphericity of 3DFSC	0.99	
Refinement		
	PfcRT	Fab-Variable Domains
Initial model used (PDB code)	-	4XMM
Model resolution (Å)		3.5
FSC threshold		0.5
Model resolution range (Å)		2.6-5.6
Map sharpening B factor (Å ²)		-28.3
Residue range	47-113 and 123-405	Heavy Chain: 4-129 Light Chain: 2-106
Model composition		
Non-hydrogen atoms	2810	1780
Protein residues	350	231
Ligands	1	0
B factors (Å²)		
Protein	81.1	74.5
Ligand	82.5	-
R.m.s. deviations		
Bond lengths (Å)	0.0074	0.0076
Bond angles (°)	1.16	1.19
Validation		
MolProbity score	1.55	1.81
Clashscore	5.91	6.63
Poor rotamers (%)	0.96	0.52
EM-Ringer Score	3.49	3.33
Ramachandran plot		
Favored (%)	96.5	93.0
Allowed (%)	3.5	6.1
Disallowed (%)	0.0	0.9

Extended Data Table 2 | Characteristics of PfCRT mutations and isoforms

a

Mutations	Origin	Associated Haplotype	Associated Phenotype	Confirmed ¹	Location	Lining the cavity	Conservation <i>Plasmodium</i> ²	Conservation Apicomplexa ³	Citation
C72S	Field	7G8	CQ-R (PPQ-S)	Yes	TM1	Yes	83%	45%	3
M74I	Field	Dd2	CQ-R (PPQ-S)	Yes	TM1	Yes	100%	55%	3
N75E	Field	Dd2	CQ-R (PPQ-S)	Yes	TM1	Yes	100%	82%	3
K76T	Field	Dd2 and 7G8	CQ-R (PPQ-S)	Yes	TM1	Yes	100%	55%	3
T93S	Field	Dd2+T93S	PPQ-R (CQ-R)	Yes	TM2	Yes	100%	64%	7,8
H97Y	Field	Dd2+H97Y	PPQ-R (CQ-R)	Yes	TM2	Yes	100%	55%	6
C101F	Lab	Dd2+C101F	PPQ-R (CQ-S)	Yes	TM2	Yes	100%	55%	12
F145I	Field	Dd2+F145I	PPQ-R (CQ-S)	Yes	TM3	Yes	67%	36%	6,27
I218F	Field	Dd2+I218F	PPQ-R (CQ-R low)	Yes	TM6	Partial	100%	55%	7,8
A220S	Field	Dd2/7G8	CQ-R (PPQ-S)	Yes	TM6	Yes	67%	55%	3
Q271E	Field	Dd2	CQ-R (PPQ-S)	Yes	TM7 - JM2	No	67%	36%	3
N326S/D	Field	Dd2/7G8	CQ-R (PPQ-S)	Yes	TM8	Yes	100%	82%	3
M343L	Field	Dd2+M343L	PPQ-R low (CQ-S)	Yes	TM9	Partial	100%	55%	6
C350R	Field	7G8+C350R	PPQ-R low (CQ-S)	Yes	TM9	Yes	100%	55%	9; this study
G353V	Field	Dd2+G353V	PPQ-R mod. (CQ-S)	Yes	TM9	Yes	100%	55%	6
I356T/L	Field	Dd2/7G8	CQ-R (PPQ-S)	Yes	TM9	Yes	100%	55%	3
R371I	Field	Dd2	CQ-R (PPQ-S)	Yes	TM9 - 10	No	67%	36%	3

b

CQ							
Isoform	Binding at pH 7.5		Binding at pH 5.5		Transport in proteoliposomes		
	K_d (μ M)	B_{max} (mol CQ/mol PfCRT)	K_d (μ M)	B_{max} (mol CQ/mol PfCRT)	1-min uptake (nmol x mg ⁻¹)	K_m (μ M)	V_{max} (nmol x mg ⁻¹ x min ⁻¹)
Control	—	—	—	—	0.12 ± 0.02	n.d.*	n.d.
7G8	0.28 ± 0.06	0.38 ± 0.02	0.21 ± 0.06	0.40 ± 0.03	3.14 ± 0.50	0.77 ± 0.15	92.20 ± 4.91
7G8+F145I	0.29 ± 0.05	0.43 ± 0.02	0.24 ± 0.06	0.39 ± 0.03	0.24 ± 0.05	0.36 ± 0.08	63.01 ± 3.57
7G8+C350R	0.30 ± 0.06	0.42 ± 0.03	0.13 ± 0.03	0.37 ± 0.02	0.80 ± 0.21	1.43 ± 0.16	17.97 ± 0.87
Dd2	0.31 ± 0.04	0.38 ± 0.01	0.20 ± 0.04	0.42 ± 0.04	3.33 ± 0.22	0.64 ± 0.07	91.37 ± 2.79
WT	0.28 ± 0.04	0.42 ± 0.02	0.28 ± 0.03	0.40 ± 0.02	0.58 ± 0.08	n.d.	n.d.

PPQ							
Isoform	Binding at pH 7.5		Binding at pH 5.5		Transport in proteoliposomes		
	K_d (μ M)	B_{max} (mol CQ/mol PfCRT)	K_d (μ M)	B_{max} (mol CQ/mol PfCRT)	1-min uptake (nmol x mg ⁻¹)	K_m (μ M)	V_{max} (nmol x mg ⁻¹ x min ⁻¹)
Control	—	—	—	—	0.09 ± 0.01	n.d.	n.d.
7G8	0.22 ± 0.04	0.40 ± 0.02	0.09 ± 0.02	0.40 ± 0.01	3.26 ± 0.33	n.d.	n.d.
7G8+F145I	0.26 ± 0.03	0.42 ± 0.01	0.22 ± 0.03	0.38 ± 0.01	1.41 ± 0.12	0.61 ± 0.15	109.30 ± 7.27
7G8+C350R	0.25 ± 0.04	0.41 ± 0.02	0.21 ± 0.03	0.39 ± 0.02	0.82 ± 0.07	0.53 ± 0.04	153.00 ± 3.10
Dd2	0.22 ± 0.03	0.39 ± 0.02	0.08 ± 0.01	0.39 ± 0.01	3.09 ± 0.41	n.d.	n.d.
WT	0.26 ± 0.04	0.40 ± 0.02	0.26 ± 0.04	0.38 ± 0.02	0.51 ± 0.04	n.d.	n.d.

a. All mutations were identified in field isolates except for C101F, which was selected under PPQ pressure in Dd2 parasites. The PfCRT 7G8 and Dd2 haplotypes are listed in Extended Data Fig. 1. The drug first listed in the 'Associated phenotype' column is thought to have been the driver of selection and the susceptibility status is indicated; the drug listed in parentheses shows the associated phenotype. **b.** Dissociation constants (K_d) and maximum binding constants (B_{max}) were determined with equilibrium dialysis using 4.2 pmol of the indicated PfCRT variant in 50 mM Tris/MES, pH 7.5 or pH 5.5, 5% glycerol and 0.1 mM TCEP. Data (mean ± s.e.m. of $n = 3$ independent experiments) were subjected to global nonlinear regression fitting in GraphPad Prism and are shown as the log-normalized mean ± s.e.m. of the fit. Transport kinetics for [³H]CQ and [³H]PPQ were determined with PfCRT variants reconstituted into proteoliposomes or with liposomes (control) using an inwardly directed pH gradient. Data (mean of $n = 2$ independent experiments of $n = 4$ technical replicates) for the 1-min uptake of [³H]CQ or [³H]PPQ (depicted in Fig. 3i) are shown for each variant tested. Data of the concentration-dependent uptake of [³H]CQ or [³H]PPQ (mean of $n = 2$ independent experiments of $n = 4$ technical replicates) shown in Fig. 3j were fitted to the Michaelis-Menten equation in GraphPad Prism to obtain the K_m and V_{max} (shown as mean ± s.e.m. of the fit).

¹These mutations were detected in *P. falciparum* parasites from areas of PPQ resistance, except for the lab-selected C101F mutation. Their contribution to PPQ resistance has been confirmed in cultured isolates or gene-edited parasite lines.

²Percentage identity of the amino acid present in *P. falciparum* wild-type (3D7) parasites, when examined across CRT orthologues from the six *Plasmodium* species listed in Extended Data Fig. 7. For *P. falciparum*, we only considered the 3D7 strain and not the CQ-resistant Dd2 and 7G8 strains.

³Percentage identity of the amino acid present in *P. falciparum* wild-type (3D7) parasites, when examined across CRT orthologues from the six *Plasmodium* species plus the five other representative apicomplexan species listed in Extended Data Fig. 7. For *P. falciparum*, we only considered the 3D7 strain and not the CQ-resistant Dd2 and 7G8 strains.

*n.d., not determined because of low transport activities.

Extended Data Table 3 | Phenotypic parasite assays with *pfcr*-modified *P. falciparum* parasite lines

	7G8 ^{7G8}	7G8 ^{7G8+F145I}	7G8 ^{7G8+C350R}	Dd2 ^{Dd2}	WT (3D7)
[³H]CQ CAR	14.5 ± 3.7	388.4 ± 38.9	406.4 ± 22.4	26.3 ± 2.9	876.7 ± 38.7
<i>n</i>	4	4	4	6	4
<i>P</i> vs 7G8 ^{7G8}	—	0.03	0.03	—	—
<i>P</i> vs Dd2 ^{Dd2}	—	—	—	—	0.01
[³H]PPQ CAR	1943.9 ± 82.2	1125.6 ± 59.6	1207.4 ± 119.8	1861.3 ± 91.8	2123.8 ± 72.2
<i>n</i>	4	4	4	4	4
<i>P</i> vs 7G8 ^{7G8}	—	0.03	0.03	—	—
<i>P</i> vs Dd2 ^{Dd2}	—	—	—	—	0.03
CQ IC₅₀	83.0 ± 7.5	7.3 ± 0.4	6.9 ± 0.3	142.9 ± 21.8	12.2 ± 0.5
<i>N</i>	4	4	4	4	4
<i>P</i> vs 7G8 ^{7G8}	—	0.03	0.03	—	—
<i>P</i> vs Dd2 ^{Dd2}	—	—	—	—	0.03
md-CQ IC₅₀	919 ± 125	47.8 ± 2.5	36.5 ± 2.0	1796 ± 199	48.6 ± 3.1
<i>N</i>	4	4	4	4	4
<i>P</i> vs 7G8 ^{7G8}	—	0.03	0.03	—	—
<i>P</i> vs Dd2 ^{Dd2}	—	—	—	—	0.03
PPQ IC₅₀	21.4 ± 1.0	28.7 ± 1.0	29.5 ± 1.6	18.3 ± 1.9	16.2 ± 1.3
<i>N</i>	4	4	4	4	4
<i>P</i> vs 7G8 ^{7G8}	—	0.03	0.03	—	—
<i>P</i> vs Dd2 ^{Dd2}	—	—	—	—	0.20
PSA 1600 nM	0.54 ± 0.20	14.7 ± 4.2	5.9 ± 0.9	0.87 ± 0.10	0.51 ± 0.17
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.04	0.04	—	—
PSA 800 nM	0.56 ± 0.17	13.1 ± 3.9	6.6 ± 0.46	0.89 ± 0.16	0.62 ± 0.16
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.04	0.04	—	—
PSA 400 nM	0.66 ± 0.20	12.7 ± 2.5	10.5 ± 1.1	1.1 ± 0.24	0.61 ± 0.19
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.04	0.04	—	—
PSA 200 nM	1.4 ± 0.2	8.6 ± 2.4	9.5 ± 1.2	1.1 ± 0.10	1.2 ± 0.54
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.04	0.04	—	—
PSA 100 nM	1.6 ± 0.3	4.0 ± 0.6	7.7 ± 2.5	1.7 ± 0.2	1.3 ± 0.35
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.04	0.04	—	—
PSA 50 nM	4.9 ± 1.3	8.5 ± 1.2	31.8 ± 9.3	7.6 ± 3.3	4.9 ± 1.1
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.04	0.04	—	—
PSA 25 nM	48.3 ± 5.5	44.5 ± 2.4	62.1 ± 9.6	35.3 ± 12.2	11.0 ± 1.7
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.79	0.25	—	—
PSA 12.5 nM	76.7 ± 5.1	62.3 ± 4.1	79.8 ± 5.5	67.7 ± 11.9	7.2 ± 1.5
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.07	0.79	—	—
PSA 6.25 nM	87.1 ± 3.5	83.4 ± 4.5	90.7 ± 5.1	83.6 ± 6.1	12.1 ± 6.8
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.07	0.79	—	—
PSA 3.125 nM	93.8 ± 3.5	95.7 ± 3.1	96.6 ± 2.4	95.9 ± 2.1	81.2 ± 19.6
<i>N</i>	5	3	3	4	2
<i>P</i> vs 7G8 ^{7G8}	—	0.07	0.79	—	—

Cellular Accumulation Ratios (CARs) of [³H]CQ and [³H]PPQ in parasitised red blood cells indicate the mean ± s.e.m., determined from 4–6 independent assays performed in duplicate. IC₅₀ values (nM) indicate the mean ± s.e.m., as determined by 4 independent assays performed in duplicate. PPQ survival assay (PSA) percentages are presented as the mean ± s.e.m. determined from 2–5 independent assays performed in duplicate. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests. *P* values are reported for comparisons with the isogenic parasite (control) lines 7G8^{7G8} and Dd2^{Dd2} (CAR), the isogenic parasite lines 7G8^{7G8} or Dd2^{Dd2} (IC₅₀) or the isogenic parasite line 7G8^{7G8} (PSA, PPQ survival assay). Cells are highlighted in grey if *P* < 0.05. md-CQ, monodesethyl-chloroquine (the active CQ metabolite); *n*, number of assays.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Leginon 3.4

Data analysis For Cryo-EM data processing, the following software were used: Appion 3.0, MotionCor2, CTFFind 4, GCTF 0.56, Relion 2.1, CislEM 1.0, CryoSPARC2, pyem 2019.
For model building, the following software were used: Coot 0.8.9, Phenix 1.13, PyMOL 2.2, Chimera 1.13, Voss Volume Voxelator 1.0.
For sequence alignment, the following software were used: OrthoMCL-DB 5, MUSCLE 3.8, ESPript 3.
For data analysis (binding and transport studies), the following software was used: Prism 7 and 8; for data display, SigmaPlot 13 was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw movie frames, micrographs, the particle stack and relevant metadata files have been deposited into EMPIAR, with accession code EMPIAR-10330. The electron density map has been deposited into EMDB, with accession code EMD-20806. The model has been deposited in the PDB, with accession code 6UKJ. All data are available in the manuscript or the supplementary materials.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Individual transport and binding experiments were routinely performed as technical replicates (n=3 or 4) to assess standard experimental variations. Parasite assays were performed in duplicate on 2 to 5 independent occasions.
Data exclusions	No data was excluded from the analyses.
Replication	Transport experiments were repeated at least 2 times, using 2 different batches of PfCRT-containing proteoliposomes. Binding studies were performed at least 3 times with protein (in nanodiscs) stemming from 2 individual preparations. Parasite data were generated on 2 to 5 independent occasions. All parasite assays conducted were successful and were included in the final analysis. Since variations of the individual proteoliposome or nanodisc batches or cell culture preparations were within the standard experimental error, all replications were considered successful and reproducible and thus included in the analyses shown in the manuscript.
Randomization	Randomization was not applicable as the quality of the preparations (nanodiscs, proteoliposomes, cell cultures) could only be assessed by performing the respective experiments.
Blinding	Blinding was not applicable to the study as no research coordinator was instated to oversee the blinding process of the different experimental approaches performed in the different participating labs of this collaborative study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit anti-GFP antibody (Invitrogen), goat anti-rabbit IgG conjugated to HRP (Abcam), horseradish peroxidase (HRP)-conjugated anti-M13 monoclonal antibody (GE Healthcare), HRP-conjugated mouse anti-human IgG F(ab') ₂ monoclonal antibody (Jackson)
Validation	Rabbit anti-GFP antibody (Invitrogen) has been validated for Western Blot with 316 publications by the manufacturer. Goat anti-rabbit IgG conjugated to HRP (Abcam) has been validated for Western Blot with 1335 publications by the manufacturer. Horseradish peroxidase (HRP)-conjugated anti-M13 monoclonal antibody (GE Healthcare) has been validated by the manufacturer for ELISA use. HRP-conjugated mouse anti-human IgG F(ab') ₂ monoclonal antibody (Jackson) has been validated for Western Blot with 7 publications by the manufacturer.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK-293 cells (ATCC), HEK-293 GnTi- cells (Invitrogen), Sf9 cells (Expression System), Plasmodium falciparum parasite lines (NIH)
Authentication	None of the cell lines has been authenticated.

Mycoplasma contamination

The cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used.

Architecture of the mycobacterial type VII secretion system

<https://doi.org/10.1038/s41586-019-1633-1>

Received: 23 December 2018

Accepted: 6 September 2019

Published online: 9 October 2019

Nikolaos Famelis^{1,2,6}, Angel Rivera-Calzada^{3,6}, Gianluca Degliesposti⁴, Maria Wingender^{1,2}, Nicole Mietrach^{1,2}, J. Mark Skehel⁴, Rafael Fernandez-Leiro³, Bettina Böttcher^{2,5}, Andreas Schlosser², Oscar Llorca^{3*} & Sebastian Geibel^{1,2*}

Host infection by pathogenic mycobacteria, such as *Mycobacterium tuberculosis*, is facilitated by virulence factors that are secreted by type VII secretion systems¹. A molecular understanding of the type VII secretion mechanism has been hampered owing to a lack of three-dimensional structures of the fully assembled secretion apparatus. Here we report the cryo-electron microscopy structure of a membrane-embedded core complex of the ESX-3/type VII secretion system from *Mycobacterium smegmatis*. The core of the ESX-3 secretion machine consists of four protein components—EccB3, EccC3, EccD3 and EccE3, in a 1:1:2:1 stoichiometry—which form two identical protomers. The EccC3 coupling protein comprises a flexible array of four ATPase domains, which are linked to the membrane through a stalk domain. The domain of unknown function (DUF) adjacent to the stalk is identified as an ATPase domain that is essential for secretion. EccB3 is predominantly periplasmatic, but a small segment crosses the membrane and contacts the stalk domain. This suggests that conformational changes in the stalk domain—triggered by substrate binding at the distal end of EccC3 and subsequent ATP hydrolysis in the DUF—could be coupled to substrate secretion to the periplasm. Our results reveal that the architecture of type VII secretion systems differs markedly from that of other known secretion machines², and provide a structural understanding of these systems that will be useful for the design of antimicrobial strategies that target bacterial virulence.

Pathogenic mycobacteria contain up to five paralogous ESX/type VII secretion systems. Three of these systems—ESX-1, ESX-3 and ESX-5—mediate the secretion of specific sets of effector proteins that have defined roles in tuberculosis¹. ESX-3 is expressed in response to iron-limiting conditions^{3,4} and has been implicated in metal homeostasis^{5–8}, inhibition of T-helper cell (CD4⁺) activation, phagosome maturation and the repair of pathogen-induced phagosomal damage^{9–11}.

ESX secretion systems feature a set of five conserved core membrane components (EccB, EccC, EccD, EccE and MycP, where Ecc denotes ESX conserved component), which mediate the secretion of the EsxA:EsxB family of virulence factors^{12–17} or of DNA¹⁸. Biochemical and structural studies of ESX-5 demonstrated that four of these components (EccB, EccC, EccD and EccE) assemble into a stable hexameric secretion pore in the cell envelope, whereas MycP—a membrane-anchored protease—is not tightly associated with the stable core^{15,19}. The coupling protein EccC recognizes effector proteins in the cytoplasm and energizes their transport²⁰.

Recent work has provided initial insights into the structural features of mycobacterial type VII secretion systems. Crystal structures of soluble domains of EccB, EccC and EccD have been determined^{20–23}. A low-resolution negative-stain electron microscopy (EM) structure of ESX-5 revealed that the type VII secretion system organizes into hexamers¹⁹;

however, the use of staining agents limited the available information to the external contour shape of the complex. Here we used single-particle cryo-EM to reveal the structure of the membrane-embedded ESX-3/type VII core complex from *M. smegmatis*, the core-complex-forming components of which (EccB3–EccE3) share 40.4–74.7% sequence identity with those of *M. tuberculosis*.

To gain insights into the molecular architecture of the type VII secretion system we expressed the ESX-3 gene cluster of *M. smegmatis*, which encodes the secreted substrates and cytosolic components alongside the components for the membrane-embedded secretion machinery (plasmid pMyNT:ESX-3; Extended Data Fig. 1a). A core complex of around 900 kDa was the most abundant species in terms of molecular mass in detergent-solubilized membrane extracts; subsequent purification revealed the ESX-3 core complex consisting of the membrane components EccB3, EccC3, EccD3 and EccE3 (Extended Data Fig. 1b, c). A minimal expression construct (pMyNT:Mini), encoding only the membrane proteins, enabled the purification of this ESX-3 core complex in high yields (Fig. 1a, b, right lanes, Extended Data Fig. 1a, d).

For cryo-EM, the sample was exchanged into the surfactant Amphipol A8-35 and was further purified by size-exclusion chromatography. In addition to the main peak at 900 kDa, we also observed larger oligomeric species that were heterogeneous in shape and size and

¹Institute for Molecular Infection Biology, Julius-Maximilians-University Würzburg, Würzburg, Germany. ²Rudolf Virchow Center for Experimental Biomedicine, Julius-Maximilians-University Würzburg, Würzburg, Germany. ³Structural Biology Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. ⁴MRC Laboratory of Molecular Biology, Cambridge, UK.

⁵Rudolf Virchow Center for Experimental Biomedicine, Electron Microscopy Facility, Julius-Maximilians-University Würzburg, Würzburg, Germany. ⁶These authors contributed equally: Nikolaos Famelis, Angel Rivera-Calzada. *e-mail: ollorca@cnio.es; sebastian.geibel@uni-wuerzburg.de

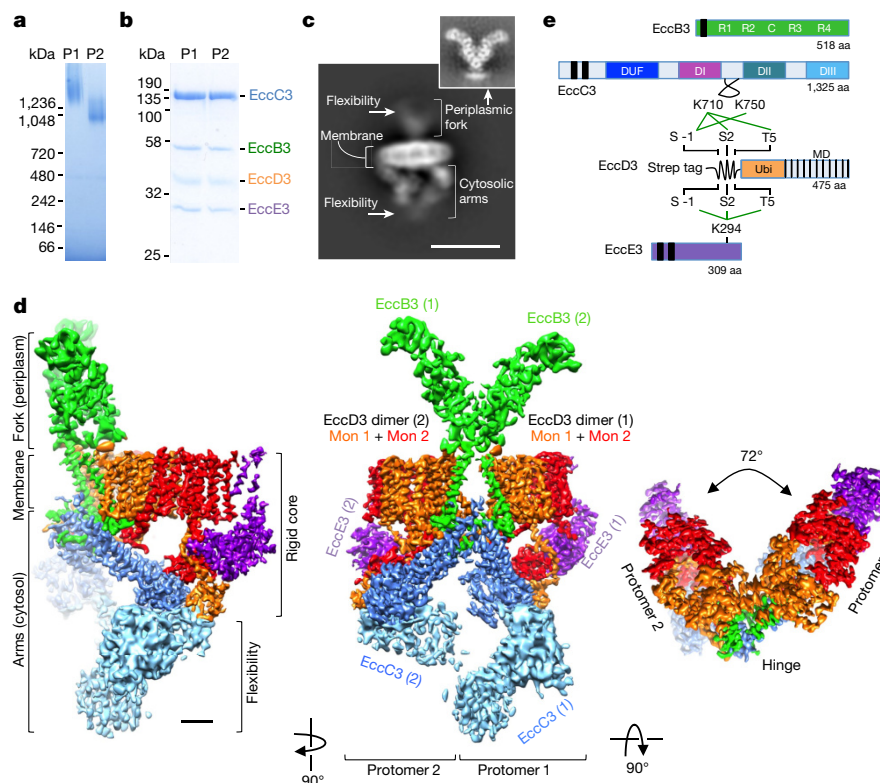


Fig. 1 | Purification and cryo-EM structure of the ESX-3 core complex. **a**, The size-exclusion chromatogram after exchange of the ESX-3 complex into the surfactant Amphipol A8-35 revealed two peaks, P1 and P2 (Extended Data Fig. 1d), which were analysed by BN-PAGE. The results are representative of more than three independent experiments. **b**, SDS-PAGE analysis of peak 1 (P1) and peak 2 (P2) of the ESX-3 core complex expressed in *M. smegmatis* from the minimal expression construct pMyNT:Mini (encoding only the membrane components of the system). The results are representative of more than three independent experiments. **c**, Representative 2D class average of the 900-kDa ESX-3 complex (peak 1). White arrows indicate flexible regions. Scale bar, 10 nm. The top inset shows the 2D average of the fork obtained after subtraction from

the images. **d**, Cryo-EM density map of the ESX-3 core complex built after merging the information of three processing strategies (Extended Data Fig. 2). In the model, EccB3 was placed in one orientation with respect to the membrane while maintaining continuity with the EccB3 TMHs inserted into the membrane. The four core components are indicated in different colours: EccB3, green; EccC3, blue and light blue for flexible regions; EccD3 dimer, monomer 1 (Mon 1) and monomer 2 (Mon 2) in orange and red, respectively; EccE3, purple. Scale bar, 2 nm. Left, side view; middle, front view; right, top view. EccB3 was removed for clarity. **e**, Schematic detailing the results of crosslinking experiments. The black bars indicate predicted transmembrane helices. MD, membrane domain. The domains of EccB3 are labelled R1, R2, C, R3 and R4.

thus were unsuitable for structure determination. However, a few two-dimensional (2D) averages that were derived from these oligomers were compatible with a complex of 25 nm in diameter—a size comparable to that of the ESX-5 hexamer¹⁹ (Fig. 1a, b, Extended Data Fig. 1d–g).

By contrast, the peak at around 900 kDa arose from a homogenous complex. The 2D class averages show a central prominent density with the distinctive shape of a membrane-spanning region. Densities on opposite sides of the membrane—which we termed ‘fork’ and ‘arms’ (Fig. 1c)—appear blurred and indicate flexibility. Independent processing of the fork revealed that this region is rigid, but has a flexible attachment to the membrane (Fig. 1c, Extended Data Fig. 1h).

The structure of the 900-kDa ESX-3 core complex was determined to a resolution of 3.8 Å, following a specific classification and refinement strategy for each region of the map²⁴ (Fig. 1d, Extended Data Figs. 1g, h, 2–5). A merged reconstruction of the complex was built using one possible orientation for the fork and one of the conformations for the arms (Extended Data Figs. 1h, 3d). The composite map shows two similar protomers interacting closely in a dimeric structure. A fork structure composed of two identical monomers is connected to the membrane (Fig. 1d). Some flexibility between the protomers was detected, and therefore the structures of each single protomer was resolved independently to improve resolution (Supplementary Tables 2, 3). The resolution of protomer 1 reached 3.7 Å, improving the definition of several regions in comparison with the structure of the dimeric core complex (Extended Data Figs. 2, 3b, 4b).

On the basis of the cryo-EM density—which shows side-chain information in most regions (Extended Data Fig. 6)—and the results of crosslinking mass spectrometry analysis (Fig. 1e), we modelled the structure for most of ESX-3 (Fig. 2). The observed crosslinks indicated the proximity of a small hydrophilic domain of EccD3 to both EccC3 and EccE3, but no crosslinks to EccB3 were found. We therefore assigned the soluble domains of EccC3, EccD3 and EccE3 to the large segment of density comprising two cytoplasmic arms, whereas the soluble domains of EccB3 reside in the smaller segment that forms the fork on the opposite side of the membrane layer (Fig. 2a, b).

Two copies of an EccB3 homology model (comprising domains R1–R4 with a putative peptidoglycan binding fold and domain C) fit into the fork (cross-correlation > 0.87) (Fig. 2b), which supports the previous proposal that EccB appends the type VII secretion system to the periplasmic peptidoglycan layer^{21,25}. The EccB3 dimer interface is formed by domains R1, R2 and C. Density is observed connecting the periplasmic domains to their respective transmembrane helices, which anchor each of the two EccB3 monomers in the membrane (Fig. 2c).

Each protomer contains 25 transmembrane helices (TMHs), comprising two copies of the EccD3 transmembrane domain (2 × 11 TMHs) for each copy of EccB3 (1 TMH) and EccE3 (2 TMHs) (Fig. 2d). Two EccD3 monomers interact in an antiparallel orientation to provide the scaffold for the ESX-3 protomer. In the membrane, TMH 11 of the two EccD3 monomers interacts with either EccB3 or EccE3. The EccD3 membrane domains are connected to their respective N-terminal ubiquitin-like

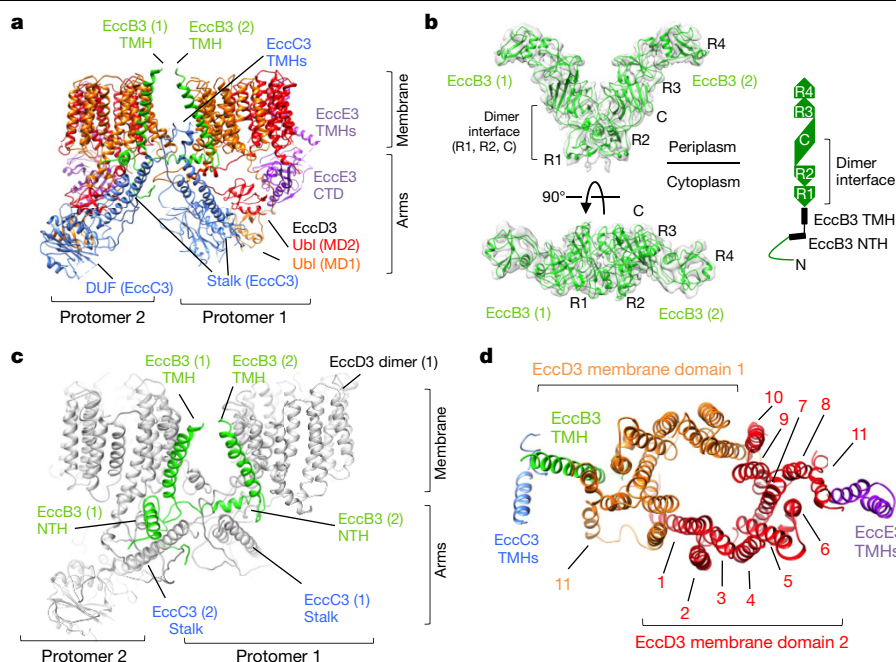


Fig. 2 | Overview of the atomic structure of the ESX-3 core complex. **a**, The rigid core of the ESX-3 dimer, excluding the flexible regions (the fork and tip of the arms). **b**, Left, fit of a homology model of EccB3 into the 4.6 Å resolution cryo-EM map. Top, side view; bottom, top view. Right, schematic of the EccB3 domain structure. **c**, Anchoring of EccB3 to the membrane of the ESX-3 dimer,

showing the TMH and the N-terminal helix (NTH). **d**, Top view of protomer 1 membrane region with the assignment of all transmembrane helices. The EccD3 dimers are shown in orange and red. The structure of protomer 1 was obtained after particle subtraction (Extended Data Figs. 2, 3), and the density for EccC3 TMH, shown in blue, is slightly better defined than in the dimer.

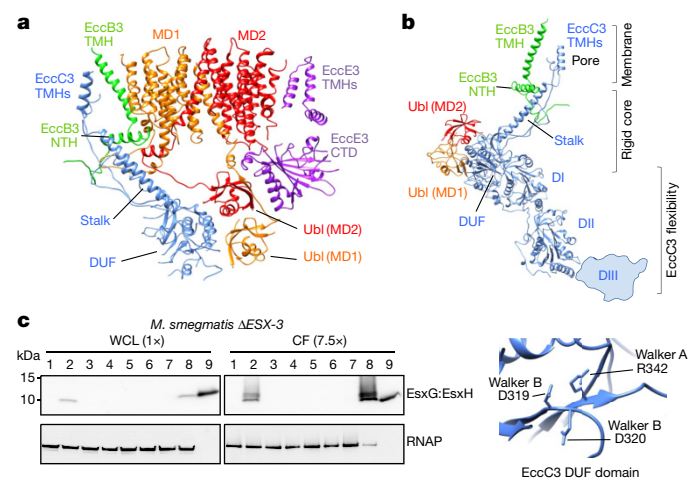


Fig. 3 | Structure and function of the ESX-3/type VII secretion system. **a**, Side view of the protomer 1 membrane region. **b**, Domain structure of EccC3, showing the stalk domain and the interaction of the DUF domain with the two Ubl domains of EccD3. **c**, The plasmid-encoded ESX-3 gene cluster (pMyNT:ESX-3) mediates secretion of the substrate heterodimer EsxG–EsxH into the culture filtrate of the transformed knockout strain *M. smegmatis* ΔESX-3. Left, whole-cell lysate (WCL) and culture filtrate (CF) of *M. smegmatis* ΔESX-3 transformants containing the following plasmids were analysed by western blot. Lane 1, pMyNT:no insert; lane 2, pMyNT:ESX-3; lane 3, pMyNT:ESX-3-ΔEccE3; lane 4, pMyNT:ESX-3-EccB3Δfork; lane 5, pMyNT:ESX-3-EccC3ΔDIII; lane 6, pMyNT:ESX-3-EccC3-D319; lane 7, pMyNT:ESX-3-EccC3-D320; lane 8, pMyNT:ESX-3-EccC3-R342T; lane 9, purified EsxG–EsxH. Loading and lysis control, RNAP (RNA polymerase β subunit; around 150 kDa). Right, a magnified view of the structure of the DUF domain, highlighting residues that are mutated in the secretion assay. The western blots are representative of two biological replicates.

(Ubl) domains through TMH1 (Fig. 3a). In the cytoplasm, the Ubl domains and their linkers to the membrane domains form a scaffold for the cytoplasmic domains of EccC3 and EccE3.

Two N-terminal TMHs insert EccC3 into the membrane. We detect partial density for these EccC3 TMHs, which could not be modelled completely (Figs. 2a, d, 3a, b, Extended Data Fig. 6b). We found that a previously unassigned stalk domain connects the EccC3 cytosolic domains to the membrane. The stalk comprises two α-helices and a short antiparallel β-sheet and makes electrostatic interactions with the EccB3 N-terminal helix, which runs almost parallel to the membrane (Figs. 2c, 3a, b). This interaction suggests that conformational changes of the stalk could be coupled to changes in EccB3. The stalk is followed by a conserved domain that is designated in literature as the domain of unknown function (DUF)²⁰. The DUF structure exhibits an ATPase fold, which is in an ATP-free state.

The flexible arms in ESX-3 could be visualized only in reconstructions that were obtained after re-centring the particles (Extended Data Figs. 2b, 3, 5). This re-centring strategy was sufficient to fit a homology model of the EccC3 DI ATPase domain within the density that is in direct contact with the DUF domain (cross-correlation > 0.87) (Fig. 3b). Classification of the particles revealed that the subsequent EccC3 ATPase domains DII and DIII are extremely flexible, and only partial density for the DII domain was detected.

Crosslinking experiments followed by mass spectrometry analysis indicated that the soluble domains of EccC3, EccD3 and EccE3 are located close to each other; the upper cytoplasmic arm was therefore assigned to component EccE3, given its localization proximal to the two Ubl domains of EccD3 and the crosslinks to these domains (Fig. 1e). EccE3 is appended to the membrane by two TMHs and forms an extensive interface with the EccD3 Ubl domain dimer in the cytoplasm (Figs. 2d, 3a). The core of the EccE3 cytoplasmic domain shows a mixed α/β-fold with a central antiparallel β-sheet surrounded by several α-helices.

Overall, the structure of the ESX-3 core complex comprises two identical protomers—each consisting of one copy of EccB3, EccC3 and EccE3 and two copies of EccD3. The EccB3 dimer is crucial for holding both

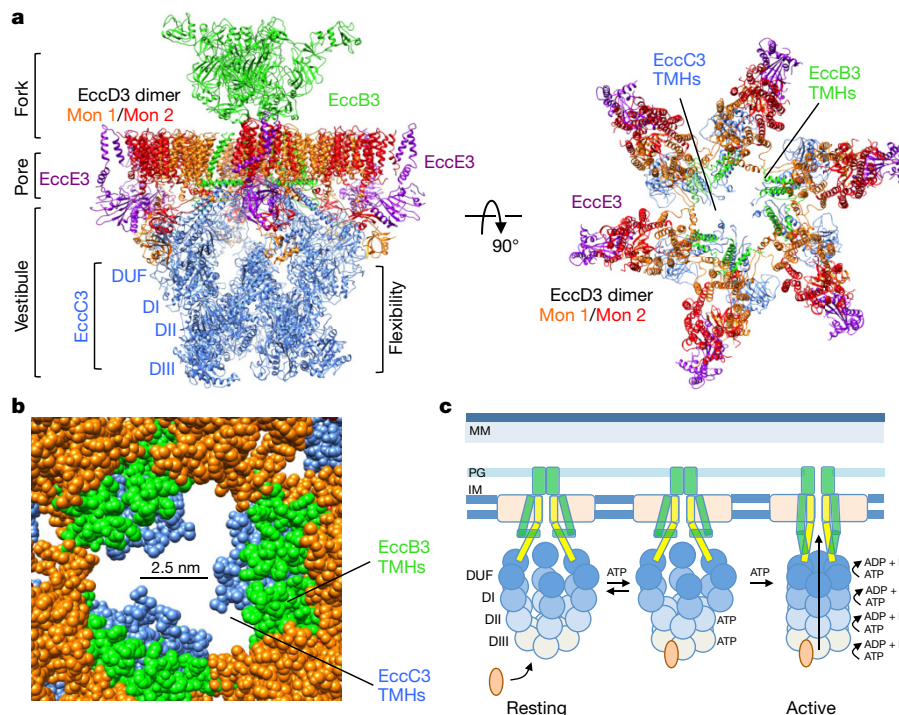


Fig. 4 | Model of the oligomeric ESX-3 secretion machine. **a**, Side and top views of the model for the ESX-3 oligomer. A vestibule-like structure is formed by the five EccC3 domains: stalk, DUF, DI, DII and DIII. **b**, Inner membrane pore and vestibule structure of the ESX-3 oligomer. The inner membrane pore is composed of EccB3 and EccC3 TMHs and the stalk domains of EccC3. **c**, Model

for the mechanism of secretion in ESX-3/type VII secretion system. The membrane pore is formed by EccC3, and contacts between EccC3 and EccB3 facilitate the coupling of ATP hydrolysis and substrate binding to changes in conformation in both proteins and the opening of the pore.

protomers together, as the periplasmic domain of each EccB3 monomer is linked to a TMH that is firmly anchored within the respective protomer (Figs. 1d, 2a, c).

We performed several experiments to validate our structure (Extended Data Fig. 7a–h). First, we confirmed that the purified membrane-embedded secretion machine is competent for secretion, and established an assay to monitor ESX-3-dependent secretion into the culture medium using antiserum against the two effector proteins, EsxG and EsxH. We used an *M. smegmatis* ESX-3 knockout strain that is defective in secretion⁵. Complementation of this strain with plasmid pMyNT:ESX-3, which encodes the ESX-3 gene cluster (Extended Data Fig. 7a), restored secretion (Fig. 3c, lanes 1, 2). Analysis of detergent-solubilized cell envelopes of the transformed knockout strain by blue native-polyacrylamide gel electrophoresis (BN-PAGE), western blot and subsequent purification confirmed the assembly of the 900-kDa ESX-3 core complex (Extended Data Fig. 7b, c).

We removed the periplasmic domain from EccB3 (using pMyNT:ESX-3) and expressed it in secretion-deficient *M. smegmatis* Δ ESX-3. We then purified the EccB3 Δ fork–EccC3–EccD3–EccE3 complex and found that the molecular mass was shifted towards that of a single protomer (Extended Data Fig. 7h). The same plasmid did not mediate the secretion of EsxG and EsxH, indicating that the periplasmic domain of EccB3 stabilizes the dimer and is essential for secretion (Fig. 3c, lane 4).

In order to investigate the function of EccE3, we removed its corresponding gene (*eccE3*) from the plasmid-encoded ESX-3 gene cluster (pMyNT:ESX-3) and expressed the remaining genes in *M. smegmatis* Δ ESX-3. Purification of the EccB3–EccC3–EccD3 subcomplex revealed that it assembled into a dimer, although it was prone to dissociation (Extended Data Fig. 7h). This plasmid did not mediate secretion (Fig. 3c, lane 3), indicating that EccE3 is a scaffold component of the ESX-3 protomer. We conclude that the stability of the ESX-3 protomers is critical for the function of the ESX-3 secretion system.

We also investigated the function of EccC3. Deletion of the DIII domain of EccC3 completely abolished secretion (Fig. 3c, lane 5), in agreement with observations in the staphylococcal type VIIb secretion system²⁶. The DUF domain contains a canonical and conserved Walker B motif, whereas the Walker A motif (336-GXXXXHRT-343) includes non-canonical residues in positions 341 (H for G) and 342 (R for K). We examined the effect of mutations in these Walker A and B motifs (Fig. 3c). D319A and D320A mutations abolished secretion, confirming that the DUF is an ATPase domain with an essential function in secretion, whereas the R342T mutation in the non-canonical Walker A motif had no effect (Fig. 3c, lanes 6–8). The domain architecture of type VII coupling proteins thus comprises a stalk domain that connects a linear array of four ATPase domains (DUF, DI, DII and DIII) to the membrane (Fig. 3b). Three out of the four ATPase domains are identified as the main constituents of the arms in the electron microscopy density (Figs. 2a, 3a, b); the density for DIII is missing in the reconstruction owing to its high degree of flexibility.

Unlike the hexameric ESX-5, we extracted ESX-3 in a dimeric state. We tested whether we could detect larger oligomers using our ‘high yield’ minimal expression construct (pMyNT:Mini, EccD3_{strep}). BN-PAGE of detergent-solubilized membrane extracts revealed species with a higher molecular mass than that of the dimers (Extended Data Fig. 7i). These species increase in amount when a crosslinking agent is used, suggesting that ESX-3 forms larger complexes in the membranes that are disrupted during our extraction protocol. As a control, large oligomers are not found in the membranes of mycobacteria that express only EccD3.

The cryo-EM structure of the ESX-3 core complex revealed that EccB3, EccC3 and EccD3 and EccE3 are present in a 1:1:2:1 stoichiometry. This is distinct from the 1:1:1:1 ratio proposed for ESX-5, which was measured by intensity-based absolute quantification mass spectrometry (iBAQ MS)¹⁹. This method was not sufficiently sensitive to determine the stoichiometry of ESX-3 (Extended Data Fig. 7f). Notably, the ratio of subunits in ESX-3 extracted from secretion-competent cells—as estimated by Oriole staining—agrees with the 1:1:2:1 stoichiometry obtained from the

cryo-EM structure of ESX-3 (Extended Data Fig. 7e). Furthermore, size-exclusion chromatography experiments found that EccD3 alone elutes as a dimer of approximately 100 kDa (2×48 kDa) (Extended Data Fig. 7g).

Using the low-resolution map of ESX-5 as a template, three ESX-3 dimers can be fitted by following the contour shape provided by ESX-5 to model the ESX-3 hexamer (Fig. 4a, b, Extended Data Fig. 8). The interior of the membrane pore has a diameter of around 25 Å and is formed by two transmembrane helices of EccC3, for which we observe partial density (Extended Data Fig. 6b). The secreted EsxG–EsxH heterodimer of the ESX-3 system, as well as the known substrate structures of other ESX systems, have a diameter of 22 Å. The secretion pore in our model is therefore sufficiently wide for the translocation of folded substrates.

On the cytoplasmic side of the proposed secretion pore, six copies of the coupling component EccC3 pack into a unique vestibule-like structure that consists of a stalk domain and four ATPase domains of EccC3 (DUF, DI, DII and DIII). The architecture of the DUF and stalk domains linked to a membrane-traversing helix that connects to a periplasmic domain is reminiscent of ABC (subfamily F) transporters (Extended Data Fig. 9). The DI, DII and DIII domains are flexible, and this flexibility increases towards the outermost layer of the vestibule, which is composed of six DIII domains. Structure–function studies on the homologous EccC coupling protein from *Thermomonospora curvata* showed that the ATPase activity of DI is modulated by its interaction with domain DII, and that peptides representing the type VII secretion signal bind to domain DIII²⁰.

The extensive contacts between the stalk domain and the two N-terminal helices in EccB3 suggest that conformational changes that are triggered by substrate binding to the DIII domain and subsequent ATP hydrolysis of the DUF could be directly transmitted to open or close the membrane pore as well as the periplasmic exit formed by EccB3 (Fig. 4c).

In summary, the detailed architecture of a type VII secretion system reveals the oligomerization of a building block comprising two tightly interconnected protomers to form a secretion pore, which is sufficient in size to enable the translocation of substrate heterodimers in their folded state (Extended Data Fig. 10). Classical antibiotics function by inhibiting essential aspects of prokaryotic biology. Pathogenic bacteria secrete virulence factors that help them to infect the host and evade the actions of the immune system. Targeting the systems responsible for secretion could provide an alternative antimicrobial strategy. Here we provide a structural and functional understanding of the architecture of type VII secretion systems, which will aid the design of strategies that target the secretion of virulence factors in *Mycobacterium* and in other bacteria for which pathogenicity depends on type VII secretion systems.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, Extended Data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1633-1>.

- Gröschel, M. I., Sayes, F., Simeone, R., Majlessi, L. & Brosch, R. ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat. Rev. Microbiol.* **14**, 677–691 (2016).
- Costa, T. R. et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015).
- Maciag, A., Piazza, A., Riccardi, G. & Milano, A. Transcriptional analysis of ESAT-6 cluster 3 in *Mycobacterium smegmatis*. *BMC Microbiol.* **9**, 48 (2009).
- Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. & Smith, I. *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect. Immun.* **70**, 3371–3381 (2002).
- Siegrist, M. S. et al. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc. Natl Acad. Sci. USA* **106**, 18792–18797 (2009).
- Serafini, A., Boldrin, F., Palù, G. & Manganello, R. Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. *J. Bacteriol.* **191**, 6340–6344 (2009).
- Tufariello, J. M. et al. Separable roles for *Mycobacterium tuberculosis* ESX-3 effectors in iron acquisition and virulence. *Proc. Natl Acad. Sci. USA* **113**, E348–E357 (2016).
- Tinaztepe, E. et al. Role of metal-dependent regulation of ESX-3 secretion in intracellular survival of *Mycobacterium tuberculosis*. *Infect. Immun.* **84**, 2255–2263 (2016).
- Mehra, A. et al. *Mycobacterium tuberculosis* type VII secreted effector EsxH targets host ESCRT to impair trafficking. *PLoS Pathog.* **9**, e1003734 (2013).
- Portal-Celhay, C. et al. *Mycobacterium tuberculosis* EsxH inhibits ESCRT-dependent CD4⁺ T-cell activation. *Nat. Microbiol.* **2**, 16232 (2016).
- Mittal, E. et al. *Mycobacterium tuberculosis* type VII secretion system effectors differentially impact the ESCRT endomembrane damage response. *MBio* **9**, e01765-18 (2018).
- Ohol, Y. M. et al. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* **7**, 210–220 (2010).
- Brodin, P. et al. Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infect. Immun.* **74**, 88–98 (2006).
- Siegrist, M. S. et al. Mycobacterial Esx-3 requires multiple components for iron acquisition. *MBio* **5**, e01073-14 (2014).
- Houben, E. N. et al. Composition of the type VII secretion system membrane complex. *Mol. Microbiol.* **86**, 472–484 (2012).
- Abdallah, A. M. et al. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol. Microbiol.* **62**, 667–679 (2006).
- van Winden, V. J. et al. Mycosins are required for the stabilization of the ESX-1 and ESX-5 type VII secretion membrane complexes. *MBio* **7**, e01471-16 (2016).
- Gray, T. A. et al. Intercellular communication and conjugation are mediated by ESX secretion systems in mycobacteria. *Science* **354**, 347–350 (2016).
- Beckham, K. S. et al. Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nat. Microbiol.* **2**, 17047 (2017).
- Rosenberg, O. S. et al. Substrates control multimerization and activation of the multi-domain ATPase motor of type VII secretion. *Cell* **161**, 501–512 (2015).
- Wagner, J. M. et al. Structures of EccB1 and EccD1 from the core complex of the mycobacterial ESX-1 type VII secretion system. *BMC Struct. Biol.* **16**, 5 (2016).
- Solomonson, M. et al. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J. Biol. Chem.* **288**, 17782–17790 (2013).
- Zhang, X. L. et al. Core component EccB1 of the *Mycobacterium tuberculosis* type VII secretion system is a periplasmic ATPase. *FASEB J.* **29**, 4804–4814 (2015).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
- McGowan, S. et al. X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proc. Natl Acad. Sci. USA* **109**, 12752–12757 (2012).
- Zoltner, M. et al. EssC: domain structures inform on the elusive translocation channel in the type VII secretion system. *Biochem. J.* **473**, 1941–1952 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Molecular biology

For the generation of the ESX-3 expression plasmids used in this study, fragments of the ESX-3 gene cluster of *M. smegmatis* mc²155 (*eccA3* to *eccE3*) were amplified by PCR (CloneAmp HiFi PCR premix, Takara), combined and cloned into the vector pMyNT using restriction-free ligation (In-Fusion HD cloning kit, Takara). The following In-Fusion primers were used for amplification (Supplementary Table 1): (i) X390/X391 (*eccA3-eccC3*); (ii) –X392/X391 (*eccB3-eccC3*); (iii) –X396/X394 (*pe5-eccE3*); (iii b) –X396/X399 (*pe5-MycP3*); (iv) –X393/X394 (*espG3-eccE3*); (v) –X395/X394 (*eccD3-eccE3*); (vi) –X390/NX391 (*eccA3-eccC3ΔD3*). A DNA fragment encoding a C-terminal His-Tag (–GSMGGSHHHHHH*) was introduced to *eccC3* using oligomers X388/X389 (vii). The backbone of the pMyNT vector was amplified using the primer pairs X386/X387 (viii) or M37/M38 (ix). In order to clone the ESX-3 gene cluster encoding a His-Tag on the 3' end of *eccC3*, fragments (i) and (vii) were inserted into the pMyNT backbone (viii) first. The resulting construct was then linearized using the restriction enzyme SnaBI followed by insertion of fragment (iii) to produce plasmid pMyNT:ESX-3 or fragment (iii b) to produce the pMyNT:ESX-3Δ*EccE3* variant. For the generation of plasmid pMyNT:ESX3i, we excluded the acetamidase promoter region of the pMyNT:ESX3 using primers X430/X426 and replaced it with the *I*deR promoter amplified with primers X427/X428.

To generate construct pMyNT:ESX-3Δ*S* (lacking all substrate genes of the ESX-3 gene cluster), fragments (i) and (vii) were cloned into the pMyNT backbone (viii). The resulting construct was digested with SnaBI followed by ligation of fragment (iv). Plasmid pMyNT:ESX-3Δ*SC* was produced by insertion of fragments (ii) and (vii) into the pMyNT backbone (viii). After digestion by SnaBI, fragment (v) was cloned into the construct.

To generate plasmid pMyNT:Mini, the multi-cloning site of the pMyNT vector, downstream of the acetamidase ribosome-binding site, was replaced with the oligonucleotide pair X397/X398. The genes *eccB3*, *mycP3* and *eccE3* were amplified from the genomic DNA of *M. smegmatis* mc²155 using In-Fusion primers X266/X267, X307/X308 and X275/X276. Gene *eccC3* encoding a 3'-terminal His-tag (–GSMGGSHHHHHH*) was amplified using In-Fusion primers X311/X312 and *eccD3* encoding an N-terminal Strep-tag II (MASWSHPQFEKGS–) was amplified using In-Fusion primers X274/X273. Genes were cloned via the following restriction sites: EcoRV for *eccB3*, HindIII for *eccC3*, SnaBI for *eccD3* (N-terminal Strep-tagII), Scal for *mycP3* and HpaI for *eccE3*.

To generate plasmid pMyNT:StrepII-EccD3 the multi-cloning site of the pMyNT vector was removed by inverse PCR using primers X42/X43. Oligonucleotides (X44/X44C) encoding the Strep-tag II and the BamHI restriction site were annealed and then inserted into the linearized vector by ligation. Gene *eccD3* was amplified with In-Fusion primers X96/X97 and cloned into the pMyNT plasmid via the BamHI restriction site.

To generate construct pMyNT:ESX-3-EccC3ΔDIII (lacking the DIII domain of EccC3), fragments (vi) and (vii) were cloned into the pMyNT backbone (viii). The resulting construct was digested with SnaBI followed by ligation of fragment (iii).

The plasmid EccB3Δ*fork* was produced by amplifying the following fragments using pMyNT:ESX-3 as template: NX1/NX2 (pMyNT-*eccA3-eccB3Δfork*), NX3/NX4 (*eccB3Δfork-eccE3*), NX5/NX6 (*eccE3-pMyNT*).

For generation of plasmids pMyNT:ESX-3-EccC3-D319A, pMyNT:ESX-3-EccC3-D320A and pMyNT:ESX-3-EccC3-R342T—containing point mutations in the Walker A or B motifs of the DUF of EccC3—the vector pMyNT was linearized by PCR first using primer pair M37/M38 (ix). Fragments (x) (*eccA3-eccC3*) and (xi) (*eccC3-eccE3*) were amplified from pMyNT:ESX-3 using In-Fusion primers M39/M40 or M47/M48,

respectively. Oligonucleotides encoding parts of the *eccC3* gene comprising point mutations D319A (M41/M42), D320A (M43/M44) or R342T (M45/M46) were annealed and ligated together with fragment (x) into the pMyNT backbone (ix). The resulting constructs were digested with SnaBI followed by ligation of fragment (xi).

Protein overexpression and purification of the ESX-3 dimer and the ESX-3 higher oligomer

M. smegmatis mc²155 was transformed with the respective plasmid and cultured in LB (Luria/Miller) medium supplemented with 0.05% (v/v) Tween 80 and 0.2% (v/v) glycerol at 150 r.p.m. Induction was performed with 0.2% (w/v) acetamide (or 200 μM 2,2'-dipyridyl in the case of pMyNT:ESX-3i) at an optical density at 600 nm (OD_{600nm}) of 0.5 and cells were grown until a final optical density of 1.4–1.6. Cells were pelleted by centrifugation and washed in 1× PBS buffer. The cell pellet was resuspended in buffer A (30 mM HEPES pH 8.0, 300 mM NaCl and 10% glycerol) supplemented with EDTA-free protease inhibitors (Roche). Cells were lysed by three passages through an Emulsiflex-C3 homogenizer (Avestin). Unlysed cells were removed by centrifugation (10 min, 10,000g). The membrane fraction was separated by ultracentrifugation (1 h, 100,000g). Membranes were solubilized in buffer A supplemented with 0.5% TeDM (Anatrace) at 4 °C for 1 h. Insoluble material was removed by ultracentrifugation (1 h, 100,000g).

The sample was supplemented with 30 mM imidazole and loaded on a HisTrap column (GE Healthcare). For wash and elution, buffer A supplemented with 0.0022% TeDM was used with 30 mM and 250 mM imidazole, respectively. TeDM was exchanged for Amphipol A8-35 at a 1:3 ratio of protein to amphipol. TeDM was removed by incubation with BioBeads (Bio-Rad) overnight. In order to separate the different oligomeric states of the complex, size-exclusion chromatography was performed using a Superose 6 Increase 10/300GL column (GE Healthcare). Fractions were analysed by SDS–PAGE and BN–PAGE followed by Colloidal Blue staining.

Protein expression and purification of the EccD3 dimer

M. smegmatis mc²155 was transformed with the pMyNT:Strep-EccD3 and cultured in LB medium supplemented with 0.05% (v/v) Tween 80 and 0.2% (v/v) glycerol at 150 r.p.m. Induction was performed with 0.2% (w/v) acetamide at an OD_{600nm} of 0.5 and cells were grown until a final optical density of 1.4–1.6 was reached. Cells were pelleted by centrifugation and washed in 1× PBS buffer. The cell pellet was resuspended in buffer A (50 mM Tris pH 8.0, 300 mM NaCl). Cells were lysed by three passages through an Emulsiflex-C3 homogenizer (Avestin). Unlysed cells were removed by centrifugation (10 min, 10,000g). The membrane fraction was separated by ultracentrifugation (1 h, 100,000g). Membranes were solubilized in buffer A supplemented with 1% *n*-dodecyl-β-D-maltopyranoside (DDM; Anatrace) at 4 °C for 1 h. Insoluble material was removed by ultracentrifugation (1 h, 100,000g).

The sample was loaded on a StrepTrap column (GE Healthcare). For wash and elution, buffer A was supplemented with 0.05% DDM and with 0.05% DDM and 2.5 mM D-desthiobiotin, respectively. The eluted protein was subjected to size-exclusion chromatography using a Superose 6 Increase 10/300GL column (GE Healthcare).

Generation of polyclonal antiserum against the substrates EsxG and EsxH

The EsxG–EsxH protein complex of *M. smegmatis* was expressed from plasmid pMAPLe3, which was provided by the Eisenberg laboratory (University of California, Los Angeles), and purified to homogeneity as described previously²⁷. Purified EsxG–EsxH was used to immunize rabbits (immunoGlobe). The polyclonal antiserum was subjected to affinity purification using a column coupled to the purified antigen.

Protein secretion assay

M. smegmatis strains were grown in 7H9-Tween medium at 37 °C to an OD_{600nm} of 0.8. For the analysis of whole-cell lysates, cells were pelleted by

centrifugation (5,000g, 10 min), washed with PBS buffer and disrupted by bead-beating with 0.1 mm glass beads (Sigma). Samples were supplemented with Novex Tricine SDS Sample Buffer (Invitrogen), boiled for 2 min and then separated on a 10–20% Novex Tricine gel for analysis by western blotting. Culture supernatants were centrifuged twice (5,000g, 4 °C, 10 min) and passed through a 0.2- μ m filter to remove cells that were not lysed. Proteins were precipitated by the addition of 10% trichloroacetic acid (overnight, 4 °C) followed by centrifugation (10,000g, 4 °C, 15 min). The precipitate was washed with ice-cold acetone, resuspended in Novex Tricine SDS Sample Buffer, boiled for 2 min and then separated on a 10–20% Novex Tricine gel. The amount of whole-cell lysates loaded correspond to 0.26 OD_{600nm} units of cells whereas the culture supernatants loaded correspond to 7.5×0.26 OD_{600nm} units of cells. Subsequent western blots were carried out according to the manufacturer's protocol (Invitrogen). Polyvinylidene fluoride membranes were incubated either using polyclonal rabbit serum against the EsxG–EsxH protein complex (approximately 4.5 μ g ml⁻¹) or a monoclonal mouse antibody against the RNA polymerase β subunit (1:2,000; BioLegend). Secondary horseradish peroxidase (HRP)-conjugated antibodies were used at the following concentrations: anti-rabbit IgG-HRP (1:10,000; Carl Roth), and anti-mouse IgG-HRP (1:10,000; Carl Roth). Proteins were visualized using the Enhanced Chemiluminescence substrate kit (Pierce) and a FujiFilm LAS-3000 luminescent image analyser. Band intensities were quantified using the ImageJ software²⁸.

Cryo-EM sample preparation

The sample was concentrated to 0.3 mg ml⁻¹ and a volume of 3.5 μ l was applied on glow-discharged holey carbon grids (Quantifoil Cu/Rh R1.2/1.3 400 mesh). Excess liquid was removed by blotting for 3 s (blot force –10) using filter paper followed by plunge freezing in liquid ethane using a FEI Vitrobot Mark IV at 100% humidity at 4 °C.

Cryo-EM data acquisition and image processing of larger complexes

For the analysis of the large complexes eluting in the void volume of the size-exclusion chromatography experiment (Extended Data Fig. 1), images were collected on a 200-kV FEI Talos Arctica electron microscope equipped with a Falcon III direct electron detector at the Spanish National Centre for Biotechnology (CNB-CSIC, Madrid). Images were recorded using the integrative mode and a nominal magnification of 75,000 \times , corresponding to 1.42 Å per pixel at the specimen level. A total dose of 50 electrons per Å² was used, fractionated in 50 frames over a 2.98-s exposure. Images were collected with a nominal defocus range of 1.4–3.2 μ m.

Video stacks (3,387) were corrected for drift (5×5 patches) and dose-weighted using MotionCor2²⁹. The contrast transfer function parameters were determined for the drift-corrected micrographs using Gctf³⁰. A first set of 2D-references were generated from manually picked particles in Relion³¹ and these were then used for subsequent automatic particle picking using Gautomatch (provided by K. Zhang, Yale University). A number of automatically selected particles (241,990) were extracted and thoroughly analysed using 2D-classification routines available in Relion and cryoSPARC³². A subset of 36,055 particles, corresponding to side views of the multimeric complex, was aligned using a 2D-reference and subsequently classified in 2D using a mask to focus classification on the density outside the membrane region. The dimensions of a representative average with density at both sides of the membrane region were measured.

Cryo-EM data acquisition of dimeric complexes

An initial dataset of the purified ESX-3 complex was collected at The Netherlands Center for Electron Nanoscopy. Processing of these data reached 7 Å resolution for the core of the dimer (Extended Data Fig. 2a). For the high-resolution structure, micrographs were collected at the Rudolf Virchow Center, Universität Würzburg using a 300 kV FEI Titan

Krios electron microscope (Extended Data Fig. 2b). Micrographs were recorded using a Falcon III direct electron detector in counting mode (Supplementary Table 2).

Videos (11,903) were recorded with an underfocus of 1.6 to 2.6 μ m, a calibrated magnification of 1.0635 Å per pixel, a dose of 1 electron per Å² per frame, accumulating a total dose of 50 electrons per Å² fractionated in 55 frames.

High-resolution cryo-EM data processing

The dataset was processed making use of tools available in Relion versions 2.1 and 3.0³¹, cryoSPARC³², cisTEM³³ and Scipion³⁴. Initially, videos coming from the different sessions were aligned and local motion was corrected using MotionCor2 with dose weighting. Micrographs that exhibited defects in the Thon rings due to excessive drift, ice contamination or astigmatism were discarded. Details can be found in Supplementary Table 2.

Particles (2,066,007) were automatically selected and submitted to several rounds of 2D-classification in order to discard bad particles using Relion (Extended Data Fig. 2b). After cleaning, 325,205 particles were used in all subsequent analyses. Next, an initial three-dimensional (3D) model was generated from the selected particles using routines available in Relion and this volume was used to further classify each of the datasets in 3D. After classification 165,412 particles were refined using Relion to generate a cryo-EM map with an estimated average resolution of 3.8 Å, using gold-standard refinement methods and the Fourier shell correlation (FSC) cut-off of 0.143 (Extended Data Fig. 4). Local resolution ranges were analysed within Relion. Refinement of one of the monomers after particle subtraction to remove the influence of the other monomer during refinement yielded 3.7 Å and further improved the resolution of some regions.

The cryo-EM map of the ESX-3 complex exhibited a lack of detail at both ends of the complex (tip of the arms and the fork), due in part to flexibility but possibly also to the increasing distance from the centre of the box. It has been observed previously that the accuracy of the alignments improves at the centre of rotation compared to more distal regions, especially in flexible and large multi-component macromolecular complexes. We applied a re-centring strategy in which the group of 325,205 particles previously selected was re-extracted and centred in either of the two distal regions of the map (Extended Data Fig. 2b), as detailed below. For the distal arm region, the selected particles centred at this region were classified using a mask that accounted for only the variability of the density corresponding to the most distal region of the molecule, and this classification was performed without further alignment of the particles. These two classes included 80,173 and 69,919 particles each, and they were further processed using gold-standard refinement methods in Relion to generate cryo-EM maps at 5.4 Å and 5.3 Å resolution (gold-standard FSC = 0.143 criterion) (Extended Data Fig. 5).

Similarly, particles were subsequently re-extracted and centred in the fork region (Extended Data Fig. 2b). Particles were then subtracted in order to remove the density that did not correspond to the fork. These particles were analysed using the *ab initio* routine available in Relion, and without assuming any symmetry. One of the maps, corresponding to 55,342 particles, exhibited features for the fork already observed in the 2D averages and preliminary processing of the data, and it was selected for further refinement using Relion and applying C2 symmetry. The map showed an estimated average resolution of 4.6 Å (gold-standard FSC = 0.143 criterion) (Extended Data Fig. 4). Differences in resolution across the reconstructed map were analysed using the local resolution routine available in Relion.

Model building and refinement

The programs RaptorX³⁵, I-Tasser³⁶ or Phyre2³⁷ were used to generate homology models of the periplasmic domain of EccB3 (amino acids (aa) 101–518), the DUF (aa 201–400), the three ATPase domains DI–DIII (aa 401–1325), Ubl domain of EccD3 (aa 6–97) based on crystal structures of

the periplasmic domain of EccB1 from *M. tuberculosis* (PDB ID: 3X3M), the DUF (PDB ID: 2IUUA), the Ubl domain of EccD1 from *M. tuberculosis* (PDB ID: 4KV2) and the three ATPase domains of EccC from *T. curvata* (PDB ID: 4NOH) as templates, respectively. As a first step, rigid-body fitting of these models into the cryo-EM density was performed using Chimera³⁸. To improve the fit of the DUF and the Ubl domains, five cycles of morphing were carried out using Phenix Refine v1.12³⁹ before parts of the structures were rebuilt using Coot⁴⁰. De novo models of the cytoplasmic segment and the transmembrane region of EccB3 (aa 9–89), the N-terminal stalk domain of EccC3 (aa 2–44 and 89–401), the transmembrane domain of EccD3 (98–471) and EccE3 (aa 2–285) were built guided by Quick2D⁴¹ and Haruspex⁴². EccC3 DI is located at the flexible arms and the resolution in one of the structures obtained after re-centring of the particles indicated clear secondary-structure elements and was sufficient to fit a homology model with high cross-correlation (>0.87).

For EccC3 domain DII, density is present only in protomer. A homology model of DII was fitted into this density as a rigid body using Chimera. Density for DIII (aa 1052–1325) was missing in both protomers. EccC3 contains 2 TMHs for which partial density was detected in the map of the ESX-3 monomer. For TMH1, aa 36–45 of the helix were built, whereas no density was present for aa 46–62. A polyalanine model of TMH2 (aa 72–90) was placed into the density and rigid-body fitting and morphing were carried out using Phenix Refine v1.12 (Extended Data Fig. 6b).

Model building and structure refinement were performed iteratively and using the information from the map of the dimer and the monomer. Five macro cycles of real-space and grouped B factor refinement of main and side chains of the entire model of the ESX-3 core complex were performed in Phenix Refine v1.12 at 3.7 Å resolution using restraints for secondary structure, rotamers as well as Ramachandran angles. Statistics for the model can be found in Supplementary Tables 2, 3.

To model the ESX-3 oligomer, a composite EM map of the EccB3:EccC3:EccD3:EccE3 complex was chosen, in which the EccB3 dimer was rotated by 10° along the two-fold symmetry axis with respect to the membrane layer permitting the fit of all three copies of this composite map into the ESX-5 EM map without clashes (EMDB-3596). To model the ATPase domains DII and DIII, the crystal structure of DI–DIII (PDB ID: 4NOH) was superimposed onto DI in the ESX-3 structure placing DII into the density adjacent to DI.

Crosslinking mass spectrometry of ESX-3

The purified ESX-3 core complex was crosslinked using the *N*-hydroxysuccinimide (NHS) ester disuccinimidyl dibutyric urea (DSBU, also known as BuUrBu). The crosslinking reactions were incubated for 45 min at room temperature at a final excess of 100-, 50- and 25-fold of DSBU with respect to the protein concentration. The reactions were quenched by adding NH₄HCO₃ to a final concentration of 50 mM and incubating for 15 min.

The crosslinked samples were freeze-dried and resuspended in 50 mM NH₄HCO₃, reduced with 10 mM dithiothreitol and alkylated with 50 mM iodoacetamide. After alkylation, proteins were digested with trypsin (Promega), at an enzyme-to-substrate ratio of 1:20, overnight at 37 °C. The samples were acidified with formic acid to a final concentration of 2% (v/v) and the peptides fractionated by peptide size-exclusion chromatography, using a Superdex Peptide 3.2/300 column (GE Healthcare) with 30% (v/v) acetonitrile/ 0.1% (v/v) trifluoroacetic acid as mobile phase at a flow rate of 50 µl min⁻¹. Fractions were collected every 2 min with elution volumes of 1.0 ml to 1.7 ml, lyophilized and resuspended in 2% (v/v) acetonitrile and 2% (v/v) formic acid.

Fractions were analysed by nanoscale capillary liquid chromatography–tandem mass spectrometry (LC–MS/MS) using an Ultimate U3000 HPLC (Thermo Scientific Dionex) to deliver a flow of approximately 300 nl min⁻¹. A C18 Acclaim PepMap100 5 µm, 100 µm × 20 mm nanoViper (Thermo Scientific Dionex), trapped the peptides before separation on a C18 Acclaim PepMap100 3 µm, 75 µm × 250 mm nanoViper (Thermo

Scientific Dionex). Peptides were eluted with a gradient of acetonitrile. The analytical column outlet was directly interfaced via a nanoflow electrospray ionization source, with a hybrid quadrupole Orbitrap mass spectrometer (Q-Exactive HF-X, Thermo Scientific). MS data were acquired in data-dependent mode. High-resolution (*R*) full scans (*R* = 120,000, *m/z* 350–2,000) were recorded in the Orbitrap and after H activation (stepped collision energy 30 ± 3) of the five most intense MS peaks, MS/MS scans (*R* = 15,000) were acquired.

For data analysis, Xcalibur raw files were converted into the MGF format through MSConvert (Proteowizard)⁴³ and used directly as input files for MeroX⁴⁴. Searches were performed against an ad hoc protein database containing the sequences of the complexes and a set of randomized decoy sequences generated by the software. The following parameters were set for the searches: maximum number of missed cleavages 3; targeted residues K, S, Y and T; minimum peptide length 5 amino acids; variable modifications: carbamidomethyl-Cys (mass shift 57.02146 Da), Met-oxidation (mass shift 15.99491 Da); BuUrBu-modification fragments: 85.05276 Da and 111.03203 Da (precision: 5 ppm MS and 10 ppm MS); false-discovery rate cut-off: 5%. Finally, each fragmentation spectrum was manually inspected and validated.

Mass determination of purified EccD3 using size-exclusion chromatography coupled to multi-angle light scattering

The experiments were performed at room temperature using a Superose 6 Increase 10/300 GL column coupled to an AKTA Purifier system (GE Healthcare) in-line with Dawn 8+ MALS and Optilab T-rEX refractive index detectors (Wyatt Technology). Purified EccD3 (100 µl, 1 mg ml⁻¹) was applied to a Superose-6 Increase size-exclusion column (10/300 GL) in a buffer containing 50 mM Tris pH 8.0, 300 mM NaCl and 0.05% (w/v) DDM. The molecular mass was determined with the ASTRA 6 software (Wyatt Technology). For analysis of the protein conjugates, we used a refractive index increment (*dn/dc*) of 0.185 ml g⁻¹ for the protein fraction and 0.133 for DDM.

Stoichiometric analyses of ESX-3 protein complexes by the iBAQ method and Oriol staining

For in-gel digestion, the excised gel slices were destained with 30% acetonitrile, shrunk with 100% acetonitrile, and dried in a vacuum concentrator. Trypsin digestion was performed overnight at 37 °C in 0.05 M NH₄HCO₃ (pH 8), using 0.1 µg of protease per slice. Peptides were extracted from the gel slices with 5% formic acid.

NanoLC–MS/MS analyses were performed on an Orbitrap Velos Pro (Thermo Scientific) equipped with a PicoView Ion Source (New Objective) and coupled to an EASY-nLC 1000 (Thermo Scientific). Peptides were loaded on capillary columns (PicoFrit, 30 cm × 150 µm i.d., New Objective) packed with ReproSil-Pur 120 C18-AQ 1.9 µm (Dr. Maisch), and separated with a 30 min linear gradient from 3% to 30% acetonitrile and 0.1% formic acid at a flow rate of 500 nl min⁻¹.

MS scans were acquired in the Orbitrap analyser with a resolution of 30,000 at *m/z* 400; MS/MS scans were acquired in the Orbitrap analyser with a resolution of 7,500 at *m/z* 400 using higher-energy collisional dissociation fragmentation with 30% normalized collision energy. A TOP5 data-dependent MS/MS method was used; dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 30 s; singly charged precursors were excluded from selection. The minimum signal threshold for precursor selection was set to 50,000. Predictive automatic gain control was used with a target value of 1e6 for MS scans and 5e4 for MS/MS scans. The lock mass option was applied for internal calibration in all runs using background ions from protonated decamethylcyclopentasiloxane (*m/z* 371.10124).

For processing of the raw data files, database searches and quantification, MaxQuant version 1.5.7.4 was used⁴⁵. The search was performed against the *M. smegmatis* reference proteome databases (UniProt) and, additionally, a database containing common contaminants. The search was performed with tryptic cleavage specificity with three

allowed miscleavages. Protein identification was under control of the false-discovery rate (<1% on protein and peptide level). In addition to MaxQuant default settings, the search was performed allowing the following variable modifications: protein N-terminal acetylation, Gln to pyro-Glu formation (N-terminal Gln), and oxidation (Met). Carbamidomethylation (C) was set as fixed modification. iBAQ intensities were used for protein quantitation. Proteins with less than two identified razor/unique peptides were dismissed.

The purified ESX-3 core complex was separated by SDS-PAGE. Gels were stained with Oriol (Bio-Rad) according to the manufacturer's instructions. Protein band intensity was measured with the Gel Doc XR+ (Bio-Rad) using the UV transilluminator. ImageJ was used for quantification²⁸.

Crosslinking of ESX-3

Crosslinking experiments were carried out using the same protocol as for the ESX-5 core complex¹⁵. In brief, mycobacterial membranes were isolated as described in the purification section, resuspended in PBS containing 250 mM sucrose, and incubated with 1 mM disuccinimidylsuberate (DSS) or dimethylsulfoxide (as a control) for 30 min on ice. The crosslinking reaction was quenched by the addition of 100 mM glycine, 10 mM Na₂HPO₄ (pH 8.5) for 30 min. The crosslinked proteins were extracted from the membrane by detergent solubilization with 0.5% DDM for 1 h on ice. Aggregates and insoluble material were removed by ultracentrifugation (100,000g, 1 h). The solubilized membrane proteins were separated by BN-PAGE using NativePAGE Novex 3–12% BisTris gels followed by western blot analysis using antibodies against the Strep-tag II (StrepMAB-Classical HRP conjugate, IBA Lifesciences) according to the manufacturer's protocol (Invitrogen).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Cryo-EM maps have been deposited in the Electron Microscopy Database under accession codes EMD-10186 (core complex dimer), EMD-10187 (protomer 1), EMD-10189 (conformation 1), EMD-10190 (conformation 2) and EMD-10188 (dimer of EccB3). The model of the ESX-3 core complex, protomer 1, and the fork have been deposited in the Protein Data Bank under PDB accession codes 6SGW, 6SGX, and 6SGY, respectively. The cryo-EM map and model for protomer 2 (discussed but not shown, because it has lower resolution than that of protomer 1) are deposited under accession codes EMD-10191 and 6SGZ. All gel source data are provided in Supplementary Fig. 1, and all other data are available from the corresponding author upon reasonable request.

27. Arbing, M. A. et al. Heterologous expression of mycobacterial Esx complexes in *Escherichia coli* for structural studies is facilitated by the use of maltose binding protein fusions. *PLoS ONE* **8**, e81753 (2013).
28. Rueden, C. T. et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).
29. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
30. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

31. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
32. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
33. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
34. de la Rosa-Trevin, J. M. et al. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
35. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
36. Yang, J. and Zhang, Y. Protein structure and function prediction using I-TASSER. *Curr. Protoc. Bioinform.* **52**, 5.8.1–5.8.15 (2015).
37. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
38. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
39. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
41. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
42. Mostosi, P., Schindelin, H., Kollmannsberger, P. & Thörn, A. Automated interpretation of cryo-EM density maps with convolutional neural networks. Preprint at <https://www.biorxiv.org/content/10.1101/644476v1> (2019).
43. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536 (2008).
44. Götz, M. et al. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *J. Am. Soc. Mass Spectrom.* **26**, 83–97 (2015).
45. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
46. Fitzpatrick, A. W. P. et al. Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump. *Nat. Microbiol.* **2**, 17070 (2017).

Acknowledgements High-resolution electron microscopy data were collected at The Netherlands Center for Electron Nanoscopy (NeCEN) with assistance from C. Diebold, at the Rudolf Virchow Center Würzburg with the assistance of C. Kraft and T. Rasmussen, and at the Centro Nacional de Biotecnología (CNB-CSIC) with the assistance of R. Arranz and F. J. Chichón. The work at the CNB-CSIC was supported by iNEXT (PID2907 to access CNB-CSIC) grant number 653706, funded by the Horizon 2020 program of the European Union. S. Lamer acquired mass spectrometry data (AG Schlosser, University of Würzburg). We thank M. Steigedal (Norwegian University of Science and Technology) and R. Manganello (University of Padova) for the *M. smegmatis* ESX-3 knockout strains; and D. Eisenberg and M. Arbing (University of California, Los Angeles) for the pMAPLe3 overexpression plasmid. Plasmid pMyNT was a gift from A. Parret and M. Wilmanns (Addgene plasmid 42191). We thank C. Fernández Rodríguez (CNIO) for suggestions during processing, and C. Kisker and H. Schindelin (Rudolf Virchow Center, Würzburg) for critical proofreading. This work was supported by the Elite Network of Bavaria (N-BM-2013-246 to S.G.) and the State Research Agency, Spanish Ministry of Science, Innovation and Universities (MCIU/AEI to O.L.) co-funded by the European Regional Development Fund (ERDF) (SAF2017-82632-P to O.L.), by projects Y2018/BIO4747 and P2018/NMT4443 from the Autonomous Region of Madrid and co-funded by the European Social Fund and the European Regional Development Fund (to O.L.). The CNIO is supported by the National Institute of Health Carlos III.

Author contributions N.F., A.R.-C., O.L. and S.G. designed the research; N.F., A.R.-C., G.D., B.B., S.G., M.W. and N.M. carried out the research; N.F. performed the biochemistry, sample preparation, grid screening and built the model together with S.G.; A.R.-C. performed the image processing and reconstructions; N.F., A.R.-C., G.D., B.B., O.L. and S.G. analysed data; G.D. and J.M.S. performed the crosslinking mass spectrometry experiments; A.S. performed the iBAQ measurements; R.F.-L. helped with electron microscopy data processing; and N.F., A.R.-C., O.L. and S.G. wrote the paper.

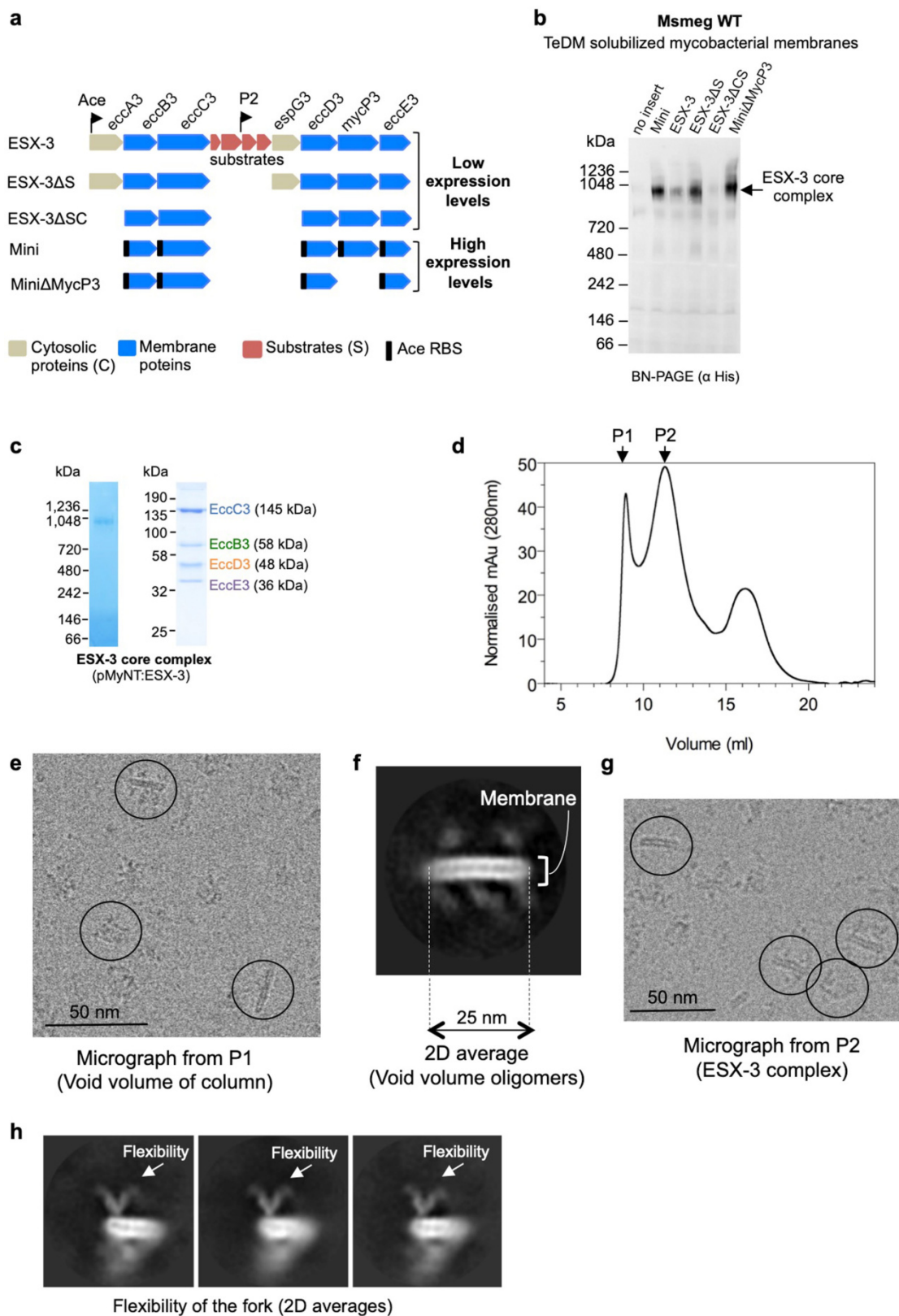
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1633-1>.

Correspondence and requests for materials should be addressed to O.L. or S.G.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

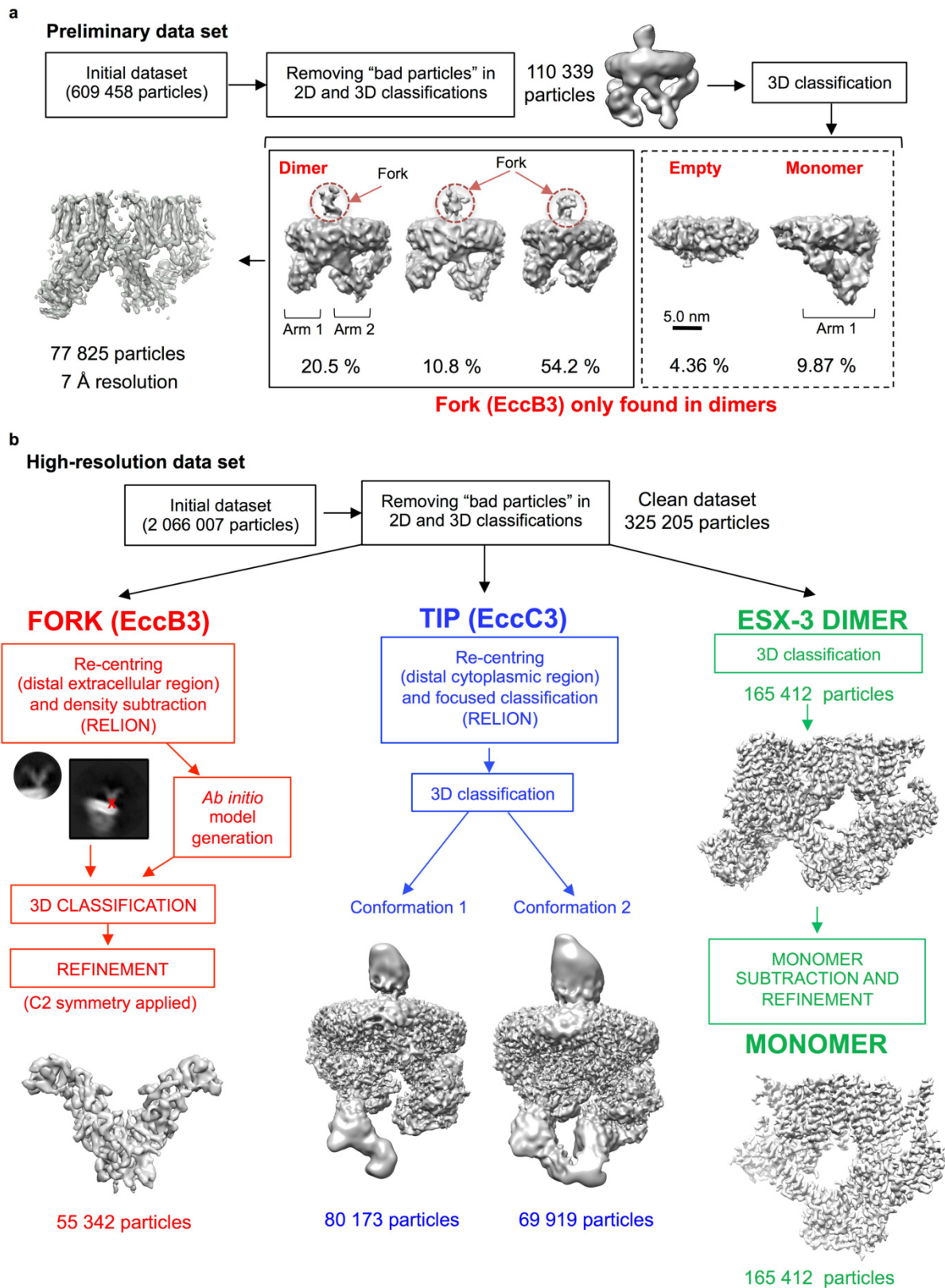


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Purification and cryo-EM of ESX-3 protein complexes.

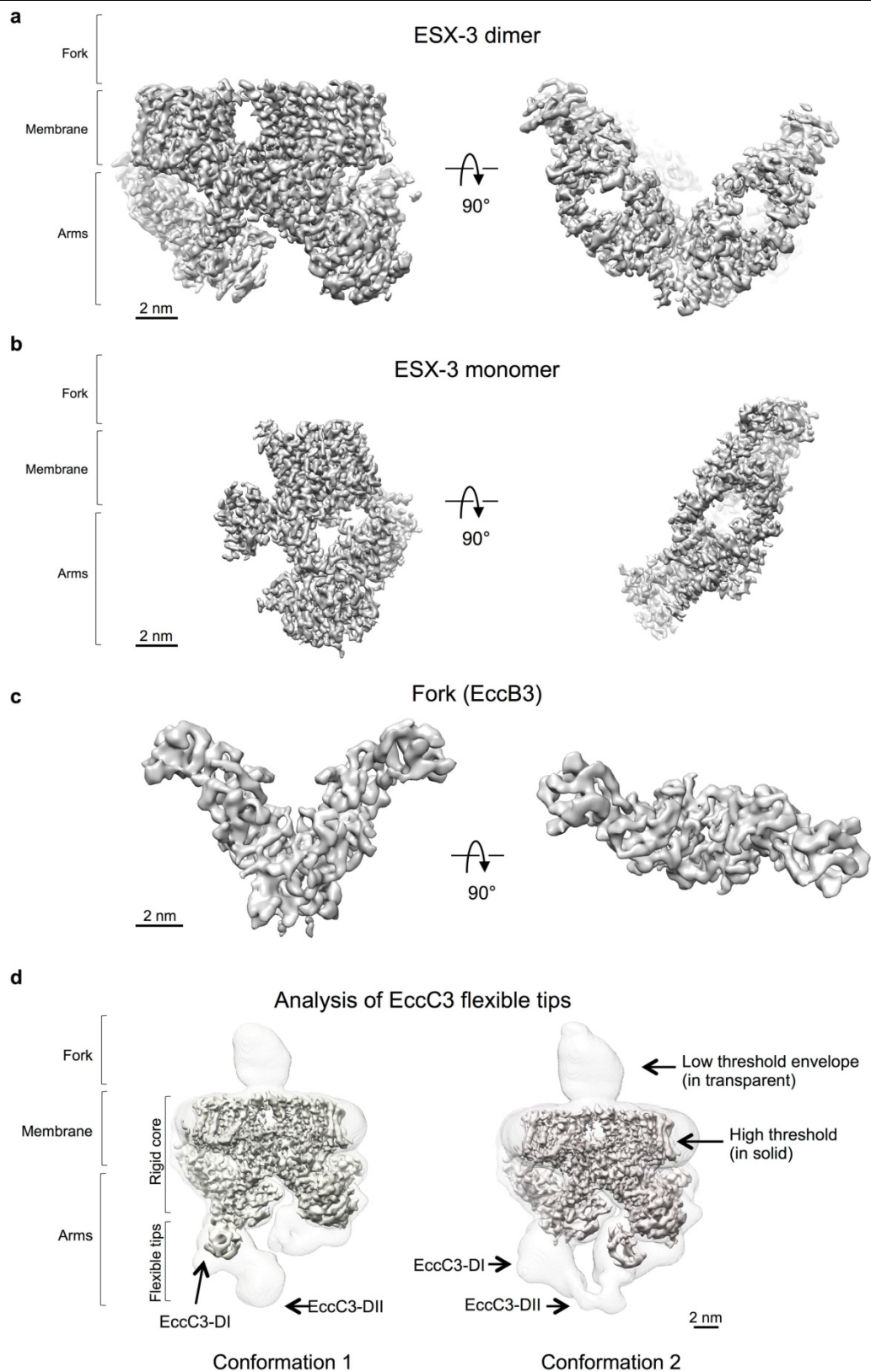
a, Overview of the constructs tested for co-expression studies. The illustrated gene cassettes were cloned into plasmid pMyNT and expressed under the control of the acetamidase promoter (Ace). Ace RBS, acetamidase ribosome-binding site. From top to bottom: the ESX-3 gene cluster (pMyNT:ESX-3), the ESX-3 gene cluster without substrate genes (pMyNT:ESX-3 Δ S), without genes encoding substrates and cytosolic proteins (pMyNT:ESX-3 Δ SC), only the ESX-3 membrane proteins (pMyNT:Mini), the membrane proteins EccB3, EccC3, EccD3, EccE3 but not MycP3 (pMyNT:Mini Δ MycP3). **b**, Western blot of TeDM-solubilized membrane extracts containing ESX-3_{EccC3-His6} complexes expressed in *M. smegmatis* mc²155 from constructs listed in **a** and separated by BN-PAGE. Anti-His₆-tag antibodies were used to detect the His-tagged ESX-3 complexes. The results represent three independent experiments. **c**, The ESX-3 core complex was expressed in wild-type *M. smegmatis* mc²155 from plasmid

pMyNT:ESX-3 encoding the entire ESX-3 gene cluster and purified; analysis by BN-PAGE (left) and SDS-PAGE (right) is shown. The results represent three independent experiments. **d**, Size-exclusion chromatogram (Superose-6 Increase 10/300 GL) after the TeDM-to-Amphipol A8-35 exchange of the ESX-3 core complex. The positions of peak 1 (P1) and peak 2 (P2) are indicated with arrows. The results represent three independent experiments. **e**, Overview micrograph of the higher oligomeric ESX-3 core complex (>1.2 MDa; P1). Some particles are highlighted by circles. Scale bar, 50 nm. **f**, Representative 2D class average of the larger ESX-3 species. The membrane region is indicated with a bracket. **g**, Overview micrograph of the 900-kDa ESX-3 core complex (P2). Some particles are highlighted by circles. Scale bar, 50 nm. **h**, Several 2D averages of the ESX-3 core complex after focused centring on the flexible fork. A flexible attachment of the fork to the membrane region can be observed.



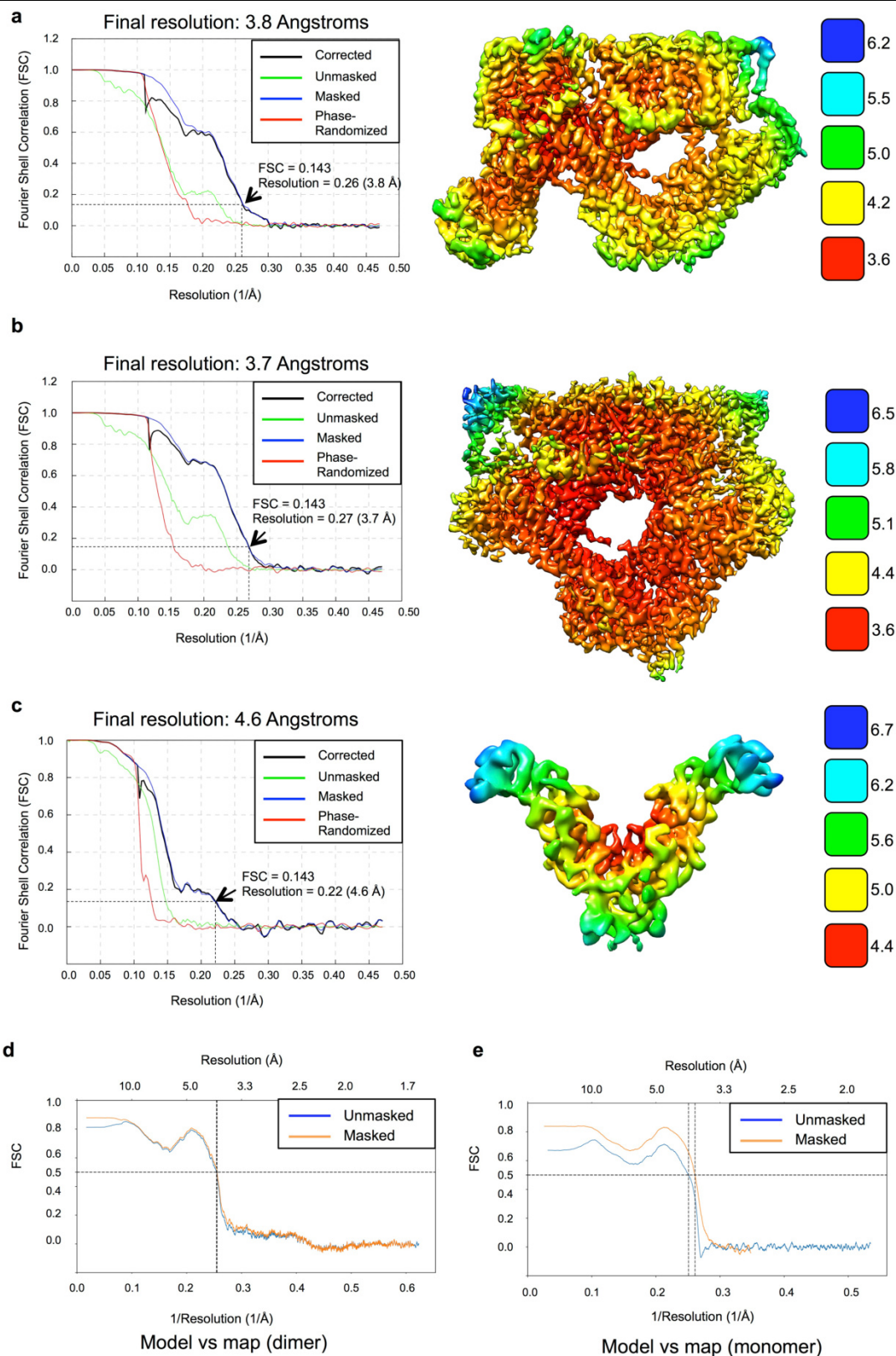
Extended Data Fig. 2 | Image processing and classification strategy for the cryo-EM data of the ESX-3 core complex. a, Tools in Relion and cryoSPARC were used to clean a preliminary dataset and classify the images. Of the particles, 85% corresponded to complexes that comprised protomers and one fork, and these were grouped and refined, whereas the single protomer did not exhibit the fork (EccB3). This analysis suggests that EccB3 is involved in dimerization. **b,** A larger

dataset was collected for high-resolution analysis. After removing bad particles, using 2D and 3D classification methods, a clean dataset was further classified and refined to the core ESX-3 dimer, and the monomer was obtained after particle subtraction. The structure of the fork was resolved after density subtraction and refinement of the new generated particles.



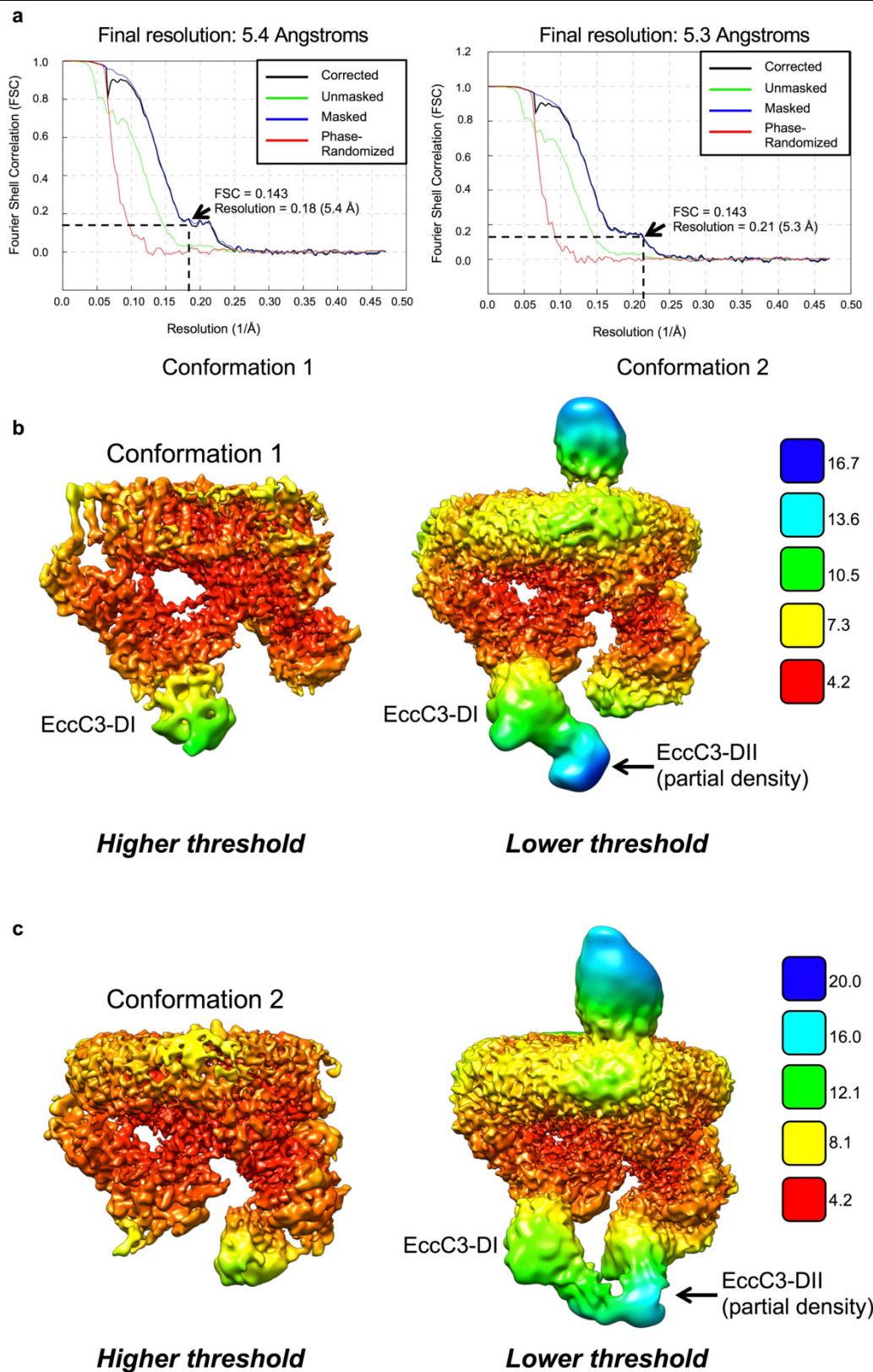
Extended Data Fig. 3 | Cryo-EM maps of the ESX-3 core complex. a–c, Two views of the ESX-3 dimer (**a**), the ESX-3 monomer (**b**) and the fork region (EccB3) (**c**). **d**, One view of each of the two conformations of the flexible tips of the arms, corresponding to EccC3, obtained after 3D classification. The density of each

cryo-EM map is rendered at a low threshold and represented as a transparent density. Within each density, the same cryo-EM map is represented at higher threshold to highlight the structural details at higher resolution in the map.



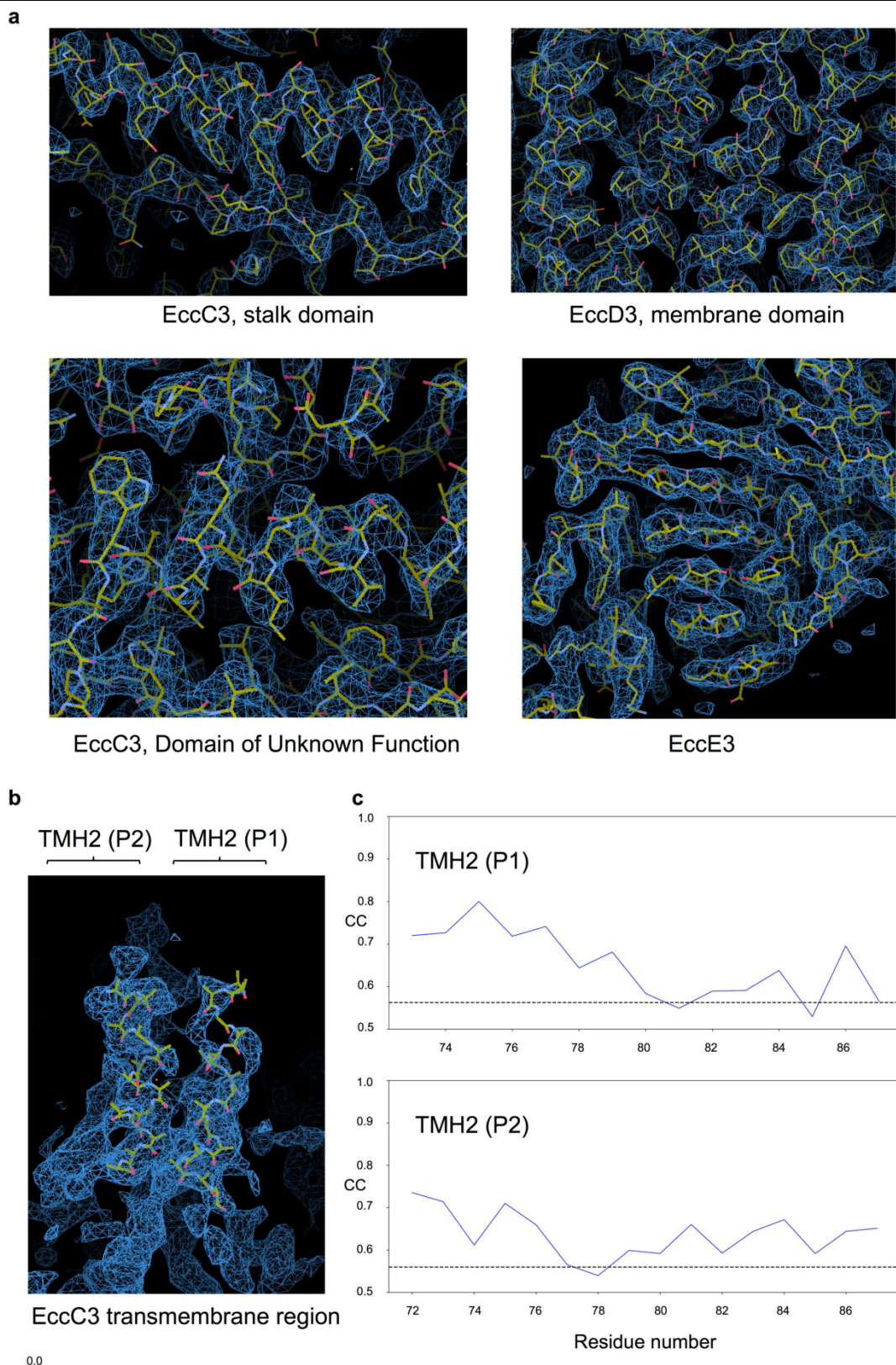
Extended Data Fig. 4 | Estimation of resolution and local resolution EM maps. a, Left, FSC plot and average resolution estimation for the core structure of the ESX-3 core complex (dimeric structure). Right, local-resolution map and

colour scale. **b,** As in **a** for the structure of protomer 1. **c,** As in **a** for the structure obtained for the fork region. **d, e,** Plot of the correlation between the model and the map for the ESX-3 dimer (**d**) and the ESX-3 protomer 1 (**e**).



Extended Data Fig. 5 | Estimation of resolution and local resolution EM maps for the flexible arms. **a**, FSC plot and estimation of average resolution for the two conformations (conformation 1 and 2) obtained for the complex after recentring and processing around the flexible tips, corresponding to EccC3. **b**, Local-resolution maps and colour scale for conformation 1, represented at

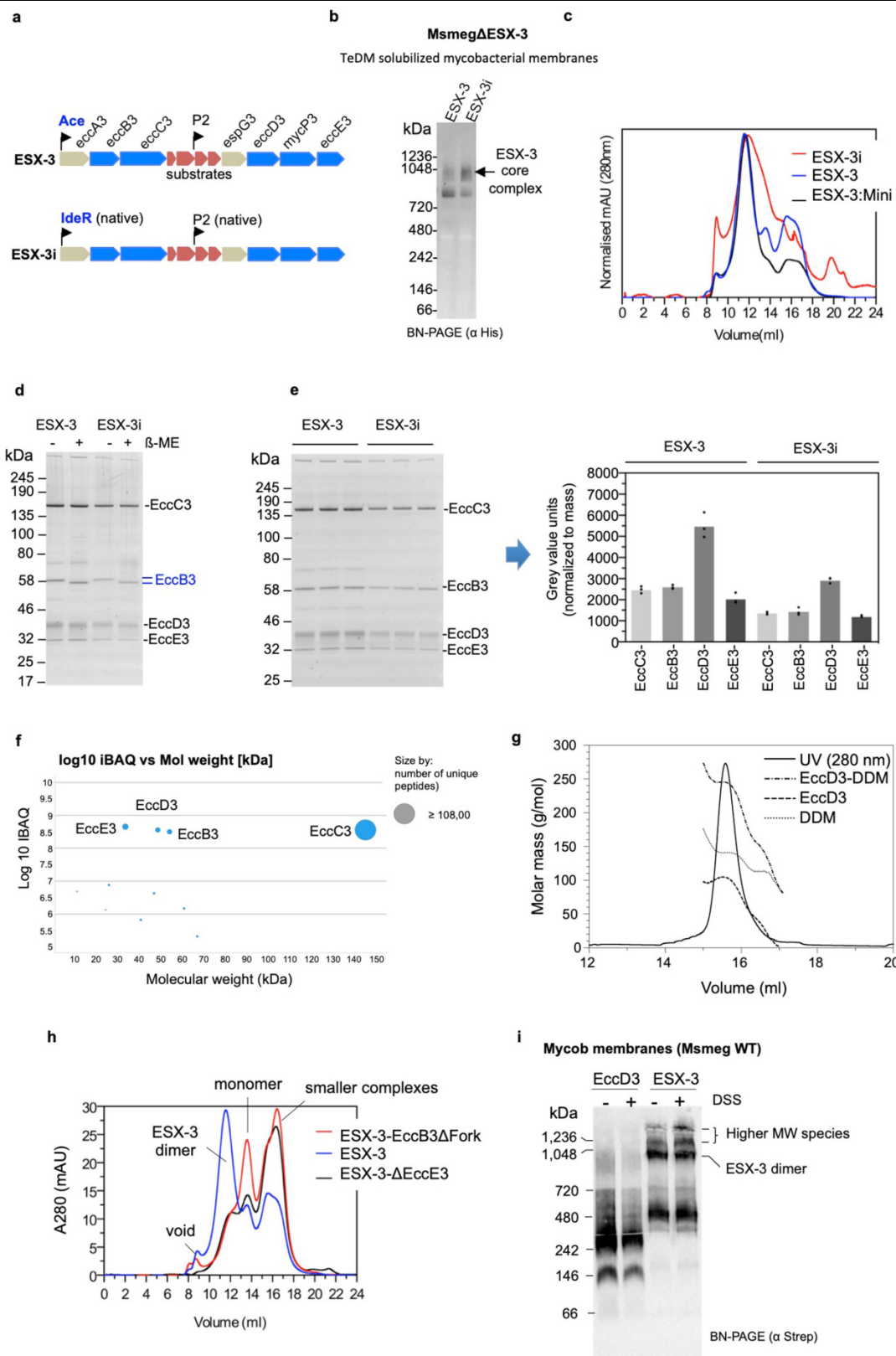
two different thresholds. At lower threshold density the whole visible tip region is represented but the higher-resolution details of EccC3 domain DI are not visible. Representations at higher threshold reveal these details at the expense of the regions of lower resolution. **c**, As in **b** for conformation 2.



Extended Data Fig. 6 | Details of the cryo-EM density and atomic models.

a, Representative regions in the cryo-EM density of protomer 1 showing details of high resolution and the atomic model for several domains in the structure. **b**, In the cryo-EM map of the ESX-3 dimer there is partial density for one TMH of EccC3 for each protomer. The two TMHs were fitted as polyalanine models into

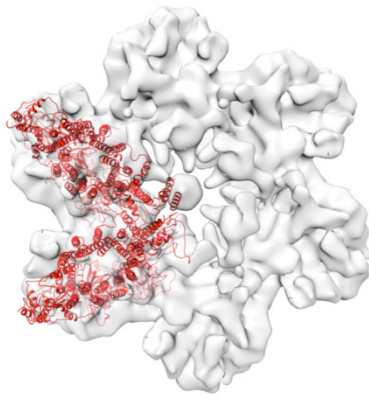
the density using Phenix Refine³⁹. **c**, Plot revealing that density for the two EccC3 TMHs is well defined only at one end of the helix. Notably, the density for one of the EccC3 TMHs is better defined in the structure of the subtracted monomer, and this is represented in the panels showing the monomer structure in Figs. 2, 3.



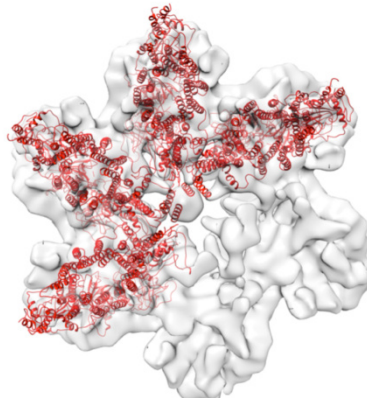
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Validation experiments. Experiments to validate that the ESX-3 core complex extracted from actively secreting mycobacteria is the same as the complex for which the structure was solved. We investigated the expression of the ESX-3 complex in the mycobacterial membrane, the stoichiometry of the extracted and purified complex and the orientation of the complex in the membrane. **a**, Overview of the constructs used in secretion assays and for the validation of the structure of the ESX-3 core complex. Plasmid pMyNT:ESX-3 encodes the ESX-3 gene cluster under acetamidase promoter control, and pMyNT:ESX-3i under control of the native IdeR promoter. The expression of the latter construct is induced in iron-depleted culture medium. **b**, BN-PAGE and western blot analyses (EccC3-His) indicate the presence of the 900-kDa ESX-3 core complex in the membrane of secretion-competent cells that contain the ESX-3 gene cluster under control of either the acetamidase promoter or the native IdeR promoter (see **a**). The gel represents two independent experiments. **c**, Comparison of size-exclusion chromatography profiles of ESX-3 core complexes purified from secretion-competent cells transformed with plasmids pMyNT:ESX-3 and pMyNT:ESX-3i (both expressed in *M. smegmatis* Δ ESX-3; see **a**), with the ESX-3 core complex purified from the minimal expression construct (pMyNT:Mini) expressed in wild-type *M. smegmatis*, which was used to determine the cryo EM structure (Extended Data Fig. 1a). The chromatograms are representative of three independent purifications. **d**, ESX-3 core complexes purified from secretion-competent *M. smegmatis* Δ ESX-3 transformed with either pMyNT:ESX-3 or pMyNT:ESX-3i, in the presence and absence of β -mercaptoethanol. A disulfide bridge is found in the EccB3 protein in complexes purified from secretion-competent cells. The two positions for EccB3 in SDS-PAGE are indicated. In the structure of the ESX-3 core complex, EccB3 is periplasmic and contributes to the assembly between protomers. We observe a different mobility for EccB3 in the ESX-3 core complex purified from secretion-competent cells when the SDS-PAGE is performed in the presence or absence of β -mercaptoethanol. This is compatible with an intramolecular disulfide bridge in EccB3, as expected in the oxidative environment of the periplasm. The gel represents two independent experiments. **e**, Determination of the stoichiometry of ESX-3 complexes in secretion-competent *M. smegmatis* Δ ESX-3 transformed with either pMyNT:ESX-3 or pMyNT:ESX-3i. Oriole staining of the SDS-PAGE bands—in order to estimate the ratio of subunits obtained—showed that EccD3 is over-represented, in agreement with the observed 1:1:2:1 stoichiometry in the structure of the ESX-3 core complex. SDS-PAGE shows that the density of the bands of EccD3 is higher than those of EccE3, despite both

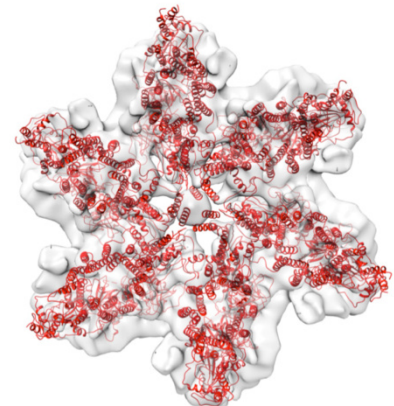
proteins having a similar molecular mass. The intensities of the bands were integrated and the relative ratio of subunits was estimated, taking into account the differences in molecular mass of the proteins. In the bar chart, the bars are the reported means and the standard deviations are shown as dots. The results indicated a stoichiometry of 1:1:2:1 (EccB3:EccC3:EccD3:EccE3). The gel shows three technical replicates and is representative of two independent experiments. **f**, Analysis of the composition of the protein complex subunit using iBAQ. The size of the dots correlates with the number of identified unique peptides. The iBAQ-MS method was, in our experiments, not sufficiently sensitive to define the stoichiometry of ESX-3. We used the ESX-3 complex from our structural studies as an internal control because the stoichiometry of the ESX-3 structure was fully determined as 1:1:2:1 at the high resolution of our cryo-EM maps. The iBAQ method estimates a 1:1:1:1 stoichiometry for the purified complex that we use to resolve the structure by cryo-EM, thus indicating that the method does not report the correct stoichiometry for ESX-3 in these experiments. When the stoichiometry of the native ESX-3 core complex in secretion-competent cells was estimated by the iBAQ method used previously for the ESX-5 core complex¹⁹, a 1:1:1:1 ratio was also obtained (data not shown). **g**, Analysis of purified EccD3 using size-exclusion chromatography (Superose-6 Increase 10/300 GL column) coupled with multi-angle light scattering. The lines show the deconvolution of the contributions of the EccD3-DDM protein-detergent complex (dot-dashed line), the protein (dashed line) and the detergent micelles (dotted line) to the total mass of the complex. This experiment demonstrates that EccD3 dimerizes as observed in structure of the ESX-3 core complex. The experiment is representative of two technical replicates. **h**, Analysis of the derivatives of the ESX-3 core complex by size-exclusion chromatography. Removing the periplasmic fork of EccB3 or EccE3 leads to dissociation towards single protomers, showing that both components are essential for the stability of the ESX-3 complex. The elution peak of the ESX-3 core complex dimer is indicated. The chromatograms are representative of two independent purifications. **i**, Isolated mycobacterial membranes from wild-type *M. smegmatis* containing the 'high yield' minimal expression construct (pMyNT:Mini) and extraction of ESX-3 core complexes after crosslinking using DSS. ESX-3 complexes were analysed by BN-PAGE and western blot analyses (EccD3-StrepII). As a control, complexes formed by EccD3 expressed alone were also analysed simultaneously. Only ESX-3 assembled into large complexes, including those of higher molecular mass (higher MW species) than the ESX-3 dimer. The gel is representative of two independent experiments.

a**TOP VIEWS**

1 dimer fitted in ESX-5



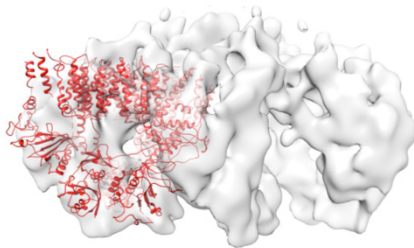
2 dimers fitted in ESX-5



3 dimers fitted in ESX-5



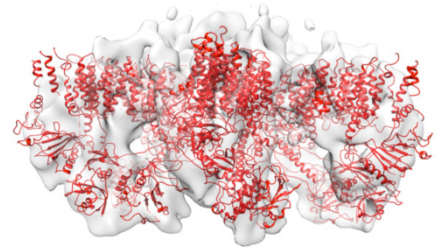
Periplasmic side



Cytosolic side

SIDE VIEWS

Periplasmic side

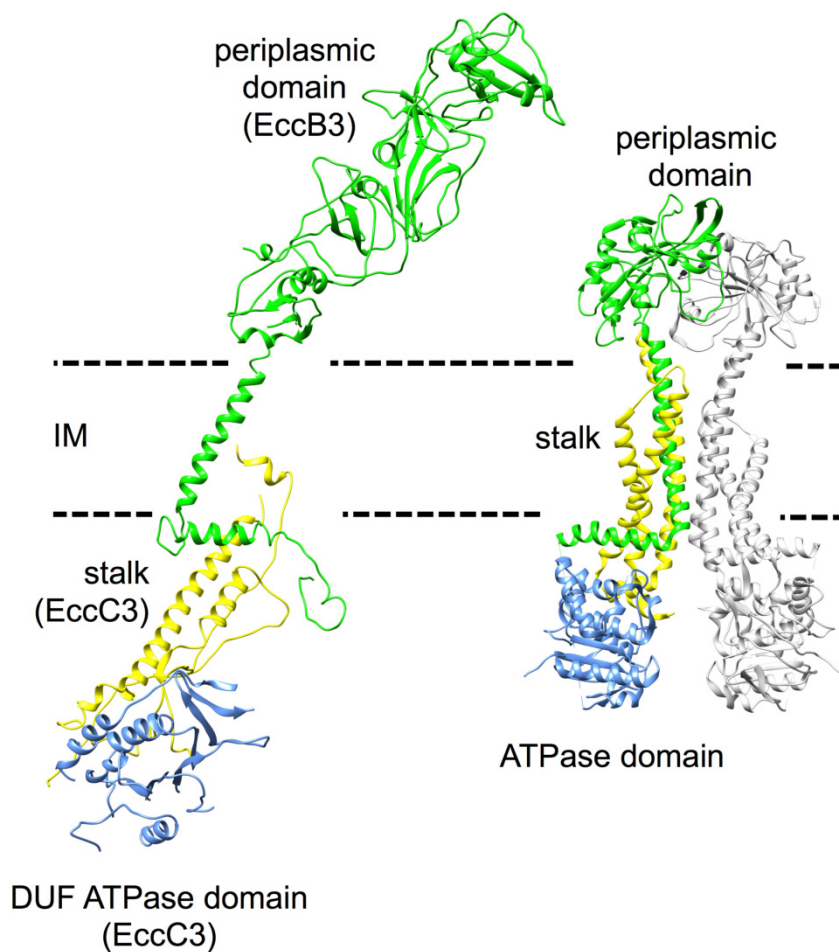


Cytosolic side

Extended Data Fig. 8 | Building the ESX-3 hexamer using the ESX-5 negative-stain structure as a template. a, Modelling of the ESX-3 secretion machine. The negative-stain map of the hexameric ESX-5 core complex (EMDB-3596) was used to generate a model for the ESX-3 secretion machine¹⁹. For this, the rigid core of the ESX-3 dimer (without the flexible arm and fork, shown in red) was fitted into ESX-5 (shown as white transparent density) using the contour information of the

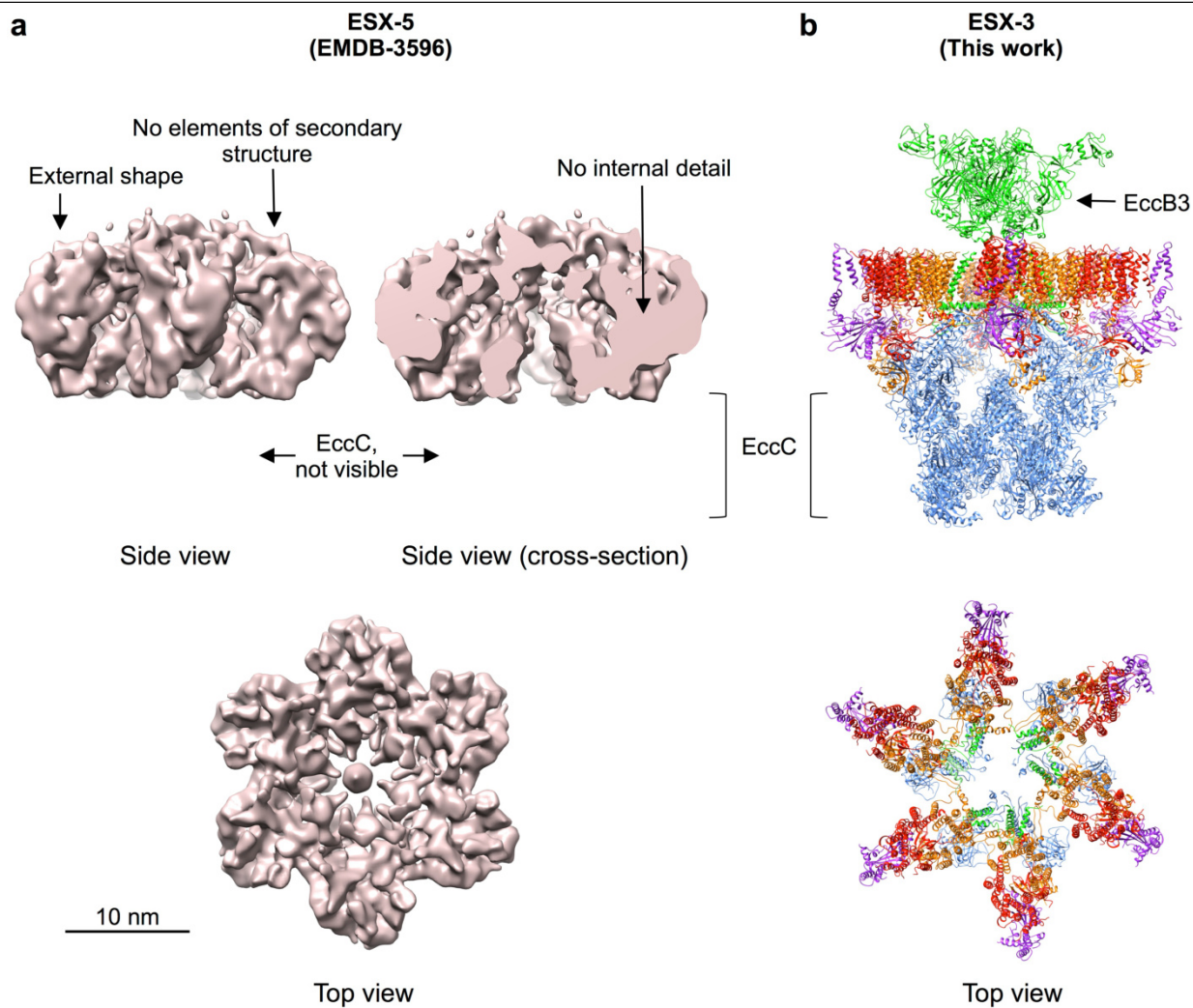
negative-stain structure. Subsequently, another two dimers were fitted in the remaining subunits to form a hexamer. The atomic structure of each ESX-3 dimer fits approximately the dimensions of ESX-5, as shown in the top and side views of ESX-5. Because the EMDB-3596 reconstruction was generated assuming a perfect six-fold symmetry, all subunits in the ESX-5 complex are identical, whereas the ESX-3 hexamer is based on a trimer of dimers.

a



Extended Data Fig. 9 | The structure of the EccC3/EccB3 contact is reminiscent of ABC type F transporters. a, Side-by-side comparison between the structure of the DUF and stalk domains of EccC3 interacting with EccB3 in the structure of the ESX-3 core complex, and the MacB ABC transporter (PDB ID:

5NIL)⁴⁶. The open (ATP unbound) state of MacB is shown. ATP binding and hydrolysis trigger long-range conformational changes in the stalk domains and the periplasmic domains of the MacB dimer, which close and open the transmembrane channel as well as the periplasmic exit.



Extended Data Fig. 10 | Comparison between the negative-stain structure of ESX-5 and the cryo-EM structure of ESX-3. a, Side and top views of the ESX-5 structure, which has been solved previously (EMDB-3596)¹⁹. The structure was obtained after applying six-fold symmetry. The use of staining agents limits the obtainable information to the external shape of the complex, while secondary

structure elements and internal details are not resolved in negative-stain structures. EccC5 and EccB5 are not visualized in the structure, possibly owing to their flexibility and the use of six-fold symmetry. **b,** Side and top views of the ESX-3 model (this work).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
<input checked="" type="checkbox"/>	<input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input checked="" type="checkbox"/>	<input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of all covariates tested
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input checked="" type="checkbox"/>	<input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input checked="" type="checkbox"/>	<input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Cryo-EM: FEI EPU 1.7.0; Western Blots: Fujifilm Imager software v3.0.
Data analysis	Cryo-EM data were processed and analyzed using RELION v2.1 and 3.0, cryoSPARC v2, cisTEM, SCIPION, LocalDeblur, MotionCor2, Gctf and Gautomatch, Chimera v1.14, Coot v0.8.9.2. COOT was used for model building. Mass spectrometry data were processed with MSConvert (Proteowizard) and MeroX software. SEC-MALS data were analyzed with ASTRA 6 software (Wyatt Technology) and Graphpad Prism v6.0h.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Cryo-EM maps corresponding to the central core region dimer, protomer 1, protomer 2, cytoplasmic distal region, and the fork have been deposited in the Electron Microscopy Database under accession codes EMD-10186, EMD-10187, EMD-10191, EMD-10189, EMD-10190 and EMD-10188 respectively and the model of the ESX-3 core complex, protomer 1, protomer 2 and the fork under accession code PDB ID 6SGW, PDB ID 6SGX, PDB ID 6SGZ, PDB ID 6SGY, respectively.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All experiments were replicated two or more times. Three proteins samples were independently purified for cryo-EM data collection and structure determination of the ESX-3 core complex. For structure determination, the sample size was determined by the selected number of particles which yielded the cryo-EM map with highest resolution.
Data exclusions	No data were excluded from the analyses.
Replication	All experiments were replicated successfully. We have provided a detailed documentation of the performed experiments as well as the procedure to derive the cryo-EM maps and the structural models. All reagents used in this study are commercially available. The cryo-EM data processing and structure modelling software used in this study is freely available.
Randomization	Samples were not allocated to groups.
Blinding	During data collection and analysis, investigators were not blinded because this procedure is not commonly applied in cryo-EM.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<p>Monoclonal Anti-polyhistidine mouse antibody peroxidase conjugate (clone HIS-1; Merck, catalog# A7058, lot #027M4768V; 1:5000).</p> <p>Monoclonal Anti-RNA polymerase mouse (RNAP) antibody (clone 8RB13; Biogend, catalog# 663903, lot# B219269; 1:2000).</p> <p>Monoclonal Anti Strep-tag II peroxidase conjugate (iba life-sciences, catalog# 2-1509-001, dilution: 1:6.000)</p> <p>Mouse IgG-HRP (GE Healthcare, catalog# NXA931, dilution: 1:10.000)</p> <p>Anti Rabbit-HRP (Carl Roth, catalog# 4750.1, lot# 4326, dilution: 1:10.000)</p> <p>Polyclonal anti-(EsxG/ EsxH) rabbit serum was generated in collaboration with immunoGlobe GmbH (Himmelstadt, Germany) as described in the method section.</p>
Validation	<p>Specificity of polyclonal anti-(EsxG/ EsxH) rabbit serum was validated by comparison of culture media of wild type <i>Mycobacterium smegmatis</i> to a <i>M. smegmatis</i> ΔESX-3 knockout strain that lacks genes <i>esxG</i> and <i>esxH</i> as well as by comparison to the purified antigens as shown in Figure 1a.</p> <p>Monoclonal anti-polyhistidine antibody peroxidase conjugate was validated by comparison to an <i>M. smegmatis</i> strain (carrying an empty pMyNT vector) that does not express His-tagged proteins (Extended Data Figure 1d).</p> <p>The specificity of the Monoclonal anti-RNAP was validated in Bergendahl V, et al. 2003 (Protein Expr. Purif. 31:155.).</p>

Author Correction: Light-entrained and brain-tuned circadian circuits regulate ILC3s and gut homeostasis

<https://doi.org/10.1038/s41586-019-1757-3>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1579-3>

Published online 18 September 2019

Cristina Godinho-Silva, Rita G. Domingues, Miguel Rendas, Bruno Raposo, Helder Ribeiro, Joaquim Alves da Silva, Ana Vieira, Rui M. Costa, Nuno L. Barbosa-Morais, Tânia Carvalho & Henrique Veiga-Fernandes

In this Letter, author Rita G. Domingues should have been associated with both 'Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal' (affiliation 1) and 'Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal' (originally affiliation 4, now renumbered as affiliation 2), instead of only affiliation 1. Affiliations 2–4 have been renumbered accordingly. The original Letter has been corrected online.

Author Correction: Chimeric peptidomimetic antibiotics against Gram-negative bacteria

<https://doi.org/10.1038/s41586-019-1810-2>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1665-6>

Published online 23 October 2019

Anatol Luther, Matthias Urfer, Michael Zahn, Maik Müller, Shuang-Yan Wang, Milon Mondal, Alessandra Vitale, Jean-Baptiste Hartmann, Timothy Sharpe, Fabio Lo Monte, Harsha Kocherla, Elizabeth Cline, Gabriella Pessi, Parthasarathi Rath, Seyed Majed Modaresi, Petra Chiquet, Sarah Stiegeler, Carolin Verbree, Tobias Remus, Michel Schmitt, Caroline Kolopp, Marie-Anne Westwood, Nicolas Desjonquères, Emile Brabet, Sophie Hell, Karen LePoupon, Annie Vermeulen, Régis Jaisson, Virginie Rithié, Grégory Upert, Alexander Lederer, Peter Zbinden, Achim Wach, Kerstin Moehle, Katja Zerbe, Hans H. Locher, Francesca Bernardini, Glenn E. Dale, Leo Eberl, Bernd Wollscheid, Sebastian Hiller, John A. Robinson & Daniel Obrecht

In Fig. 1 of this Article, the label for residue 24 in peptide 4 should have been 'Thr' instead of 'Dab', and in peptide 5, the labels for the 'N' and 'C' termini should have been reversed (with the C terminus next to Dab, and the N terminus next to Trp). This figure has been corrected online.

Publisher Correction: Imaging work and dissipation in the quantum Hall state in graphene

<https://doi.org/10.1038/s41586-019-1823-x>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1704-3>

Published online 21 October 2019

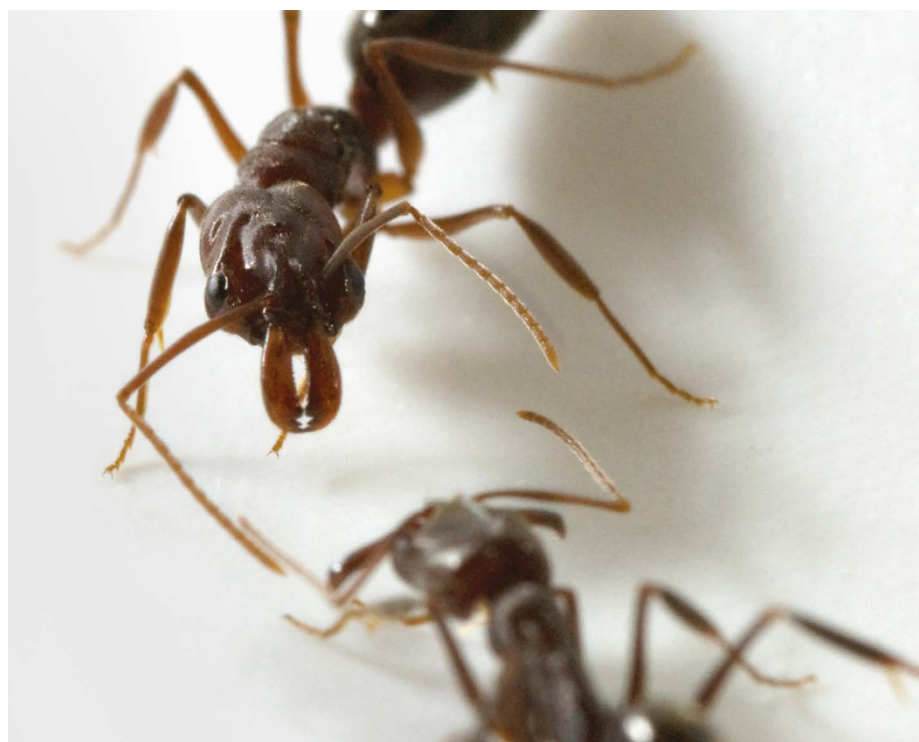
**A. Marguerite, J. Birkbeck, A. Aharon-Steinberg, D. Halbertal,
K. Bagani, I. Marcus, Y. Myasoedov, A. K. Geim, D. J. Perello & E. Zeldov**

In this Article, owing to an error during the production process, the dashed lines defining the physical edge of graphene in Fig. 2a and b were missing. This has been corrected online.

Work

Your
story

Send your careers story
to: naturecareerseditor@nature.com



Adrian Smith's footage of trap-jaw ants (genus *Odontomachus*) was a surprise hit on YouTube.

BITES, CAMERA, ACTION...

Filming my work with ants has changed my science-communication tactics. **By Adrian Smith**

Nine months ago, in my research laboratory at the North Carolina Museum of Natural Science, in Raleigh, I filmed myself being bitten by an ant. It wasn't a bite from an ordinary, everyday ant. It was my main study organism, a trap-jaw ant of the genus *Odontomachus*, with jaws that snap shut faster than almost any other recorded animal movement. It's so fast that visualizing it requires filming at a minimum of 60,000 frames per second. When I show high-speed videos of these ants, and talk about them, inevitably I'm met with the question: "Would it hurt if they snapped against you?" That's a question that was answered almost immediately for me when I started working with them eight years ago: no. They'd snap their tiny jaws at my hands and bounce off, nearly unnoticed, while

I scavenged through excavated nest soil in the field or cleaned their nest boxes in the lab.

But that question inspired me to expand my approach to communicating science. Until then, I had focused on imparting the ends of my scientific pursuits – the research results – but had overlooked opportunities to get across the fundamental, and often more exciting, aspects of my research: the initial experiences and observations. I realized that I had more research stories to tell,

"The force of the strike deflects off my unharmed finger back to the ant, sending her flying."

beyond just the final results. So, I switched on my high-speed camera and stuck my finger in front of a trap-jaw ant.

Snapping the snap

The four-and-a-half-minute video shows, in super-slow motion and in close-up, macro detail, a trap-jaw ant snapping against the tip of my finger (see go.nature.com/2da5dy). The force of the strike deflects off my unharmed finger back to the ant, sending her flying off the platform she was standing on. It's a shot that no one had filmed before, to my knowledge, contextualizing the snap of these ants on a human scale. The remainder of the video is me in the lab adding a "how" and "why" narrative to the footage. It's not a video presenting a scientific result: it's a video about what it's like to experience and observe these ants as a researcher. I uploaded it to my lab's YouTube channel and, to my surprise, a production company making a show for the US television channel Animal Planet saw it and asked to license the story. The full video will be shown on the channel's *How Do Animals Do That?* programme. The licensing fee that came to my lab funded a summer resident assistant's stipend for a graduate student.

Before all this happened, I assumed that the value of a researcher's science-communication output depended on published research findings – not on personal experiences. Its value was not monetary; rather, it was assessed by the impact of its results, the level of press exposure, or the extent to which it generated excitement and appreciation for science. Most of my output had been videos summarizing articles that I and a few colleagues had published in scientific journals. I'd attach these to an institutional press release – and if mass-media outlets picked up the release and associated media, I'd have an opportunity to communicate my work to a broad audience.

Tip of the iceberg

These products worked, and I did find a lot of value in communicating research in this way (I wrote a column about this process last year: A. A. Smith *Nature* **556**, 397–398; 2018). But it was a struggle to release more than two or three videos a year, because they were dependent on research projects progressing through peer review and on to publication. I was communicating the peer-reviewed tip of a scientific iceberg. The rest of my work, some of the most interesting parts, remained untold and hidden from view.

At its core, my job as a research scientist is to try to see and interpret the world in a

way that has not been attempted before. As a behavioural ecologist, I begin most of my research projects by trying to make and document an original observation about the life of an organism. But many of my observations end up never being documented in my peer-reviewed manuscripts. They might, for example, raise questions well beyond my area of expertise; or it could be (especially with observations that I was lucky to make) that trying to replicate them in a scientific study is just too time-consuming or costly. But I am starting to realize that none of those reasons prevents me from turning those observations into interesting stories that can drive public interest in science, and that have value beyond the rather narrow world of academic behavioural ecology.

Take my recent research experience with ant stingers. We know little about the mechanics of an ant sting. For instance, no one has ever filmed how venom is delivered out of the stinger. So, I spent a week perfecting camera techniques to gather slow-motion video of these microscopic bits of ant anatomy in action. What I recorded were the first detailed videos of venom being pumped from the stinger. From this footage, I could assess what was and wasn't measurable, and what expertise I'd need from a collaborator to move this project forwards. Beyond that, however, this bit of observational science was a fascinating story that I had to share.

Stinger zinger

I collated, edited and uploaded a 3.5-minute video showing the footage and explaining why I thought it was original and interesting. When it debuted, I also published a short Twitter thread (see go.nature.com/35kjfy) summarizing the video and providing some excerpts. The thread went viral, trending on the site's news feed and gathering more than two million impressions on the platform. That week, I spoke about the footage on two local news broadcasts; and several mass-media news outlets, including *Science*, published articles embedding video of the stinger footage that I had just captured. On YouTube, the footage garnered more than 250,000 views, and messages appeared in my inbox from people who had seen the footage and wanted to work with me.

I'm not sure whether these ant-stinger observations will lead to a peer-reviewed, published piece of science. I hope they do, but it wouldn't be unusual if this project failed to progress to that stage. What I do know is that expanding my science-communication efforts to pass on more than just the end products of my scholarly work has added a lot of value to what I produce as a scientist.

Adrian Smith is head of the Evolutionary Biology & Behavior Research lab at the North Carolina Museum of Natural Sciences and a research assistant professor in biology at North Carolina State University in Raleigh.

Mike Flannigan Fire researcher

Mike Flannigan is a professor with the Department of Renewable Resources at the University of Alberta, Canada. Here he speaks to *Nature* about his work and the boom in wildfires worldwide.

How did you get into wildfire research?

I've always been interested in fire. When I turned one year old, I burnt my finger on the candle of my birthday cake because I was fascinated by the flame. And I've been tasting fire ever since. You might call me a pyrophile.

I did an undergraduate physics degree at the University of Manitoba in Canada and, after a one-year course in meteorology, I worked as a weather forecaster for the national weather service from 1979 to 1981. I then joined the Canadian Forest Service (CFS) as a physical scientist. After completing a PhD in plant sciences at the University of Cambridge, UK, in 1993, I returned to work at CFS for 31 years. Now I'm a fire professor at Alberta.

What have you learnt from your research?

In Canada, the area burnt by wildfires has doubled since 1970. They previously burnt around one million hectares a year. Since 1990, we have been averaging more than 2.5 million hectares a year, and this is because of climate change. In the western United States, the area burnt has actually quadrupled since the 1970s, in large part because of climate change.



A wildfire burns near Ashcroft, Canada.

Q&A



XINLI CAI

Has the increase created more opportunities for wildfire researchers?

Yes. Canada, for example, is developing a Can\$50-million (US\$37.6-million) satellite for detecting and monitoring wildfires. As a research team at Alberta, we hope that we will soon get Can\$5 million in state-funded fire research. Things have changed in the field. We are actually getting funding now.

What kinds of scientist can enter the field?

Fire is a fairly broad topic. Meteorologists, physicists, atmospheric scientists, engineers, forestry specialists, geographers, computer-science programmers – you name it – could get into the field. But fire also has an operational side. Canada spends Can\$800 million a year directly on fire management, so there are a lot of professional positions in fire management. Federal, provincial and state conservation agencies hire research scientists, technical analysts and field technicians. Forest services, parks and land-management agencies are also recruiting. If the topic of your thesis includes something directly or indirectly related to fire, then you have an opportunity to get a job.

What should the public know about fire?

Fire is a global issue. Around the world, 350 million to 450 million hectares of land – close to the size of India – burn every year. In many regions, we are seeing fires that are more difficult to extinguish, and we will probably continue to experience more catastrophic events and, unfortunately, more loss of life. As an international community, we have to learn to live with fire and to address it globally. Because people start some 90% of wildfires, we need to be careful and observant. If you see a wildfire, report it right away to the authorities. We should also consider volunteerism. Programmes such as Firewise USA and FireSmart in Canada help homeowners and communities to protect themselves against fire through education, guidelines and more. Such organizations are always looking for volunteers.

Interview by Stav Dimitropoulos

This interview has been edited for length and clarity.



Where I work Mona Nemer

Walking re-energizes me when I have difficult decisions to make. I have a lot of them: as chief science adviser to the Canadian government, my role involves convening experts from multiple disciplines to offer guidance on the complex issues facing society, from renewable energies to the hazards of microplastics.

Near our office, there is a path along the Ottawa River that goes by the Parliament and Supreme Court buildings and makes a circuit, with Ottawa on one side and Gatineau, Quebec, on the other. Its natural beauty helps me to reflect on complex issues and on unexpected scientific findings.

In May 2018, the Minister of Fisheries, Oceans and the Canadian Coast Guard asked me to convene an advisory panel on how to use science better in decisions regarding aquaculture sites and regulations. I assembled a team, but there were collisions at the start because participants embraced opposing views on the value and safety of fish farming. After two days of intense discussions, I wasn't sure we could reach

a consensus. I took a walk by the water and asked myself: "What is it we all have in common and that we all value?" I realized that everybody values the environment and wants what's best for Canadians.

During my walk, I developed a strategy to establish our common ground by asking team members to identify what was most valuable to them. We eventually recommended, in part, that the government use best scientific practices, including local and Indigenous knowledge, adopt an open-science approach and prioritize regional ecosystem differences in making decisions.

The plants, flowers, trees and shrubs along the path, and the changing colours of the leaves, are chemistry and biology in action. You also come to the locks of the historic Rideau Canal. This confluence of science and technology reminds me how far we've come – and of the importance of science for providing solutions for society.

Mona Nemer is a molecular biologist and Chief Science Advisor of Canada. **Interview by Kendall Powell.**

Photographed for *Nature*
by Rémi Thériault.

nature

index

Materials science

REMASTERING MATTER

Materials for a smarter world

Hot properties

The search for new industries

Energetic performers

Solutions from early-career stars

Bounty hunters

Scavengers capture untapped power

Materials science

Editorial Catherine Armitage, Bec Crew, Rebecca Dargie, Gemma Conroy, David Payne **Analysis** Bo Wu, Catherine Cheung **Art & design** Madeline Hutchinson, Denis Mallet, Wojtek Urbanek **Production** Kay Lewis, Ian Pope, Nick Bruni, Bob Edenbach, Joern Ishikawa **Marketing & PR** Stacy Best Ruel, Angelica Sarne **Sales & partner content** Sabrina Ma, Jennie Xu, Pinky Zhang, Alex Yu, Sicong Wang, Helen Hill, Chris Gilloch, Lisa Truong **Publishing** Rebecca Jones, Richard Hughes, David Swinbanks

Nature Index 2019 Materials science a supplement to *Nature*, is produced by Nature Research, the flagship science portfolio of Springer Nature. This publication is based on data from the Nature Index, a Nature Research database, with a website maintained and made freely available at natureindex.com.

Nature editorial offices
The Campus, 4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0)20 7833 4000
Fax: +44 (0)20 7843 4596/7

Customer services
To advertise with the Nature Index, please visit natureindex.com or email clientservicesfeedback@nature.com
Copyright © 2019 Springer Nature Limited, part of Springer Nature.

All rights reserved.

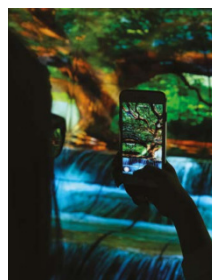
As scientists race to find low-carbon technologies to reduce greenhouse-gas emissions, energy is a hot research topic in materials science. It is obvious, then, that the search for better batteries is a focus for some emerging stars in materials science (see page S26). We also explore the field of energy harvesting, to find that everything old is new again: the cellulose fibres in wood, cotton and paper, turn out to have a cellular structure that can be treated to harvest energy from low-heat sources such as the body, opening the possibility that 'old' materials might power the smartphones and other small devices of the future (see page S38).

It's one thing to invent, discover or repurpose a material in a lab and publish it in a scientific journal, but transforming it into a commercially successful product, or the basis of an entirely new industry, is a huge accomplishment. The journey can take decades, be fraught with pitfalls and dependent on patient, deep-pocketed commercial partners. Yet the rewards can be immense (see page S20). Although we look to materials science innovations in the fight to solve climate change, they are, at best, one shaft of light in what must be a rainbow spectrum of solutions.

As government-backed initiatives in the United States and China compete to resurface 'lost' research, the next big discoveries could be made by machine-learning algorithms and data mining. China has emerged as a formidable force, although a spread of countries is still evident among the top-performing institutions (see page S42). No fewer than 43 of the 50 institutions that clocked the biggest rises in materials science research output in the Nature Index journals from 2015 to 2018 are from China (see page S34). In 2018, the annual growth in China's materials science Share in the Nature Index was 15.8%; the US's Share in the same year fell by 10.3%.

Readers will note that this supplement introduces this change in our metrics terminology: our signature metric, Fractional Count (FC) is now called Share. It is a measure of an institution's contribution to articles in the 82 journals tracked by the index, based on the proportion of its affiliated authors on an article's author list, with each author deemed to have contributed equally. The new term for our former Article Count (AC), which gives an institution, or country/region a score of 1 for every article on which one or more of its affiliated authors appear, is simply Count. You can find out more at go.nature.com/indexcount

Catherine Armitage
Chief editor



On the cover

A smartphone snaps an image of curved OLED TVs at an electronics trade fair.
Cover image: Sean Gallup/Getty Images

Contents

S20 A matter of time and risk

In a material world, scientific discoveries can revolutionize industries, but the road is long.

S26 People of substance

Early-career scientists are expanding the physical world in their search to urgent problems.

S34 Scaling up

A graphic overview shows China's rapid rise in materials science output.

S38 Energy harvesters pick up power

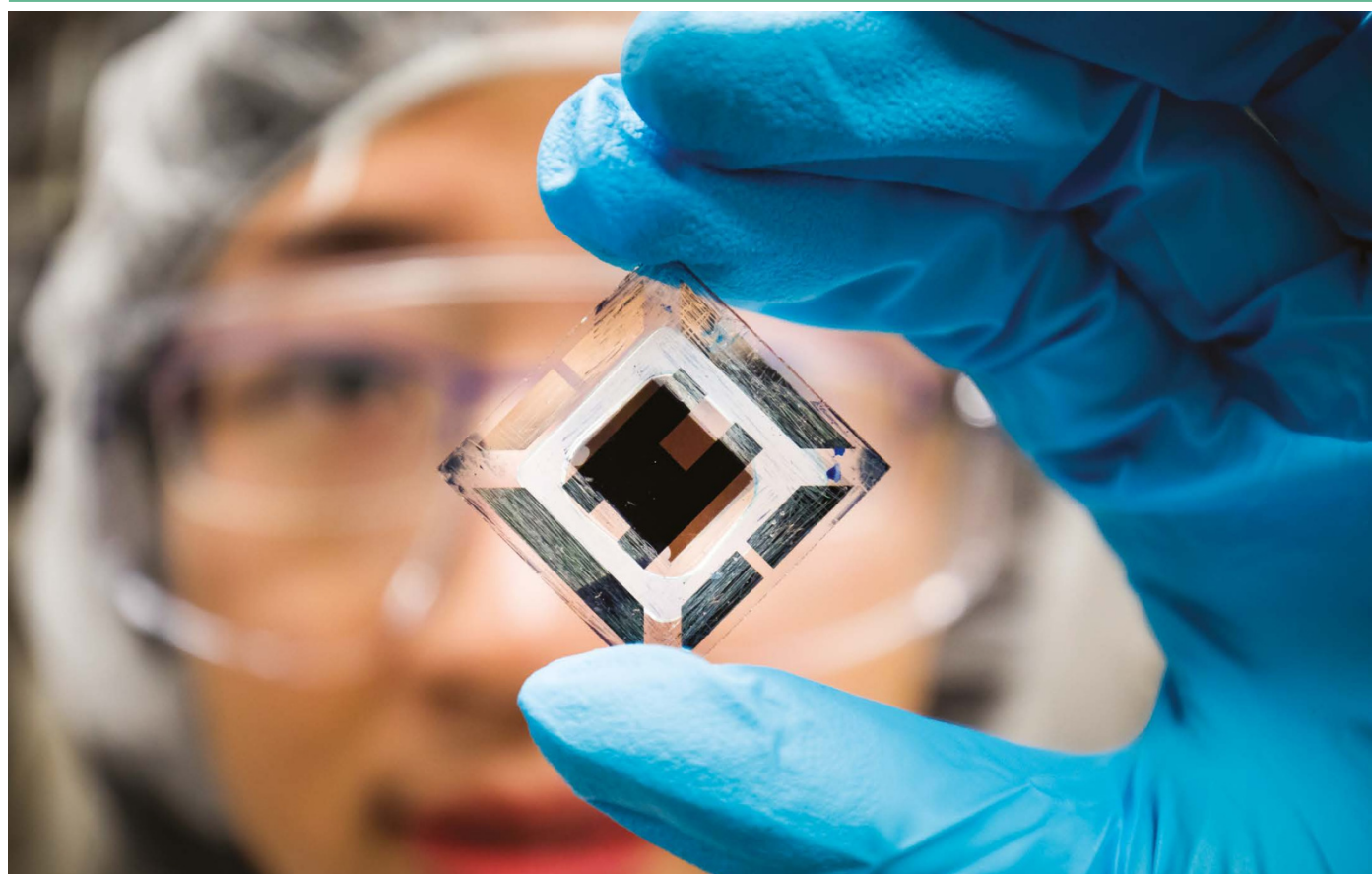
New technologies are channelling incidental energy into practical uses.

S42 The tables

How the world's institutions stack up in materials science.

A matter of time and risk

In a material world, scientific discoveries can revolutionize industries, but the road is long. **By Neil Savage**



JOSEPH XU, MICHIGAN ENGINEERING COMMUNICATIONS & MARKETING

Renewable energy conversion rates are improving through breakthroughs in technologies such as organic solar cells.

In the 1960s, the smart advice to a young person in search of a lucrative career was ‘plastics’, as canonized in the 1967 film, *The Graduate*. Today, the advice might be ‘OLEDs’ or ‘nanomaterials’. In another 20 years, it might be ‘graphene’ or ‘topological materials’.

The hunt for new, commercially viable materials with highly sought-after properties is one of the most competitive pursuits in scientific research. It offers clear applications, strong paths to market and is among the most interdisciplinary of the physical-sciences research fields. Even so, relatively few discoveries, such as organic light-emitting diodes (OLEDs) and, to a lesser extent, graphene, rise above the

throng of nascent materials; many languish in the scientific literature.

The stakes are high: a new material in the right place and at the right time can launch a multibillion-dollar industry. Consider what happened for Charles Goodyear. When he mixed sulfur with natural rubber and invented vulcanized rubber in 1839, his discovery gave rise to a multitude of new rubber products with unprecedented tensile strength and elasticity. It advanced the automobile industry, freeing motorists from unwieldy iron wheels. With the smoother ride offered by rubber tyres came the impetus to create the first gas-powered cars. More recently, global industries producing smartphones, wearable medical devices,

renewable energy and new aeroplanes have emerged from materials science research.

Artificial intelligence might prove to be the Charles Goodyear of our era. Machine-learning techniques are being used to comb through published research to unearth materials already discovered but forgotten, and to predict the properties of new materials that don’t yet exist. This replaces the trial-and-error approaches the field has relied on. According to James Warren, a physicist at the US National Institute of Standards and Technology in Gaithersburg, Maryland, in a very dynamic field it’s one of the most exciting developments in recent years.

Warren is director of the Materials Genome Initiative (MGI), a multi-agency project set up

by the US government in 2011. The database of materials and their properties is open and available to researchers, to help move those materials from journal pages into potential products, says Warren. “The whole point of MGI is literally to bridge that gap,” he says.

In just a few years, he says, it may be possible to tell a computer that you want to build a battery with a certain storage capacity, lifetime and cost, and have it suggest the optimal combination of materials to make it happen.

A long road

It can take decades for a new material to have an impact. The worldwide market for OLEDs reached US\$19.45 billion in 2017, and is expected to grow to \$81.76 billion by 2026. But, the OLEDs in your TV are a world away from the first experiments in the late 1950s to generate light from organic materials.

Stephen Forrest, a physicist at the University of Michigan, Ann Arbor, who invented phosphorescent OLEDs, is well aware of that. In 1994, New Jersey-based company, Universal Display Corporation, was founded to develop displays based on OLEDs. The company is now valued on the NASDAQ at more than US\$9 billion, and has more than 5,000 issued and pending patents worldwide, with offices in the United States, China, South Korea, Japan and the United Kingdom. “The technologies that we had [at launch] were not particularly compelling to the market,” Forrest says. “Things happened both within the marketplace and within our lab and it sort of exploded, and now we have a \$20-billion global display market based on OLEDs, and every one of them use our materials.”

Forrest has helped establish several companies based on his research, covering areas such as organic solar cells, which are lighter and more flexible than their silicon counterparts, and indium gallium arsenide detectors for infrared imaging. But it’s not just a matter of creating a new type of material and setting up shop. “You’ve got to be solving a real problem that people have an interest in,” he says. “You can’t invent markets through work in the laboratory.”

Companies often say it takes at least 10 years for a material to move from the laboratory to the market, says Zhenan Bao, a chemical engineer at Stanford University in California who works with nanomaterials and flexible electronics, including organic semiconductors and carbon-based circuits. In 2010, Bao’s lab spun out C3 Nano, a company that makes a flexible, transparent electrode based on silver nanowires. It’s being used in some foldable displays and mobile phones available in China, she says, and may soon spread elsewhere. Since she was at Bell Labs (now owned by Nokia) 20 years ago, she has been working

on foldable displays that could fit more phone screen in a given space.

Bao’s focus has broadened to developing wearable sensors that adhere to the skin and track pulse and other health indicators (see ‘Wearing well’, page S22), as well as electronic skin, a flexible, transparent material made of a self-healing silicone polymer with a conduc-

“It’s not just a matter of creating a new material and setting up shop. You’ve got to be solving a real problem.”

tive layer. It can sense pressure and send an electrical signal to the brain through an electronic artificial nerve.

Other materials researchers are developing electronic skin, including Benjamin Tee at the National University of Singapore and Shengqiang Ren at the State University of New York at Buffalo.

Graphene could be an important component of the necessary flexible electronics for

Bao’s electronic skins. Its unique properties have excited scientists since the material was first isolated in 2004. A one-atom-thick layer of carbon, graphene is stronger than diamond, has the highest electron mobility of any known material, and is more than 97% transparent to wavelengths of light from the ultraviolet to the far infrared. Those properties could come in handy, but so far, no one has come up with an application that has had much of an impact in the marketplace. Grand View Research estimates the global market for graphene was only \$43 million in 2017, mainly in applications such as protective coatings for flexible electronics.

The discovery of graphene has, however, opened up research into other 2D materials – atomically thin layers of an element or compound. These may be essential to overcoming the looming end of Moore’s law, when the transistors that power computers physically can’t get any smaller. Forrest says material that can move electrons more rapidly than silicon will be necessary to keep increasing computing power.

Quantum computing, a fledgling technology based on the laws of quantum mechanics, might benefit from an emerging class called

TOP 5 CORPORATE–ACADEMIC COLLABORATIONS*

Rank	Institution 1	Institution 2	Countries	Bilateral CS	Count
1	Samsung Group	Sungkyunkwan University (SKKU)	South Korea	25.89	32
2	Center for Nanotechnology GmbH	University of Münster (WWU)	Germany	20.99	41
3	Samsung Group	Stanford University	South Korea/US	13.92	17
4	Samsung Group	Korea Advanced Institute of Science and Technology (KAIST)	South Korea	11.91	17
5	IBM Corporation	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland/US	11.12	16

*Date range for data is 2015–18

TOP 5 CORPORATE INSTITUTIONS*

Rank	Institution	Country	Share	Count	Materials Science (%)
1	IBM Corporation	United States	112.92	223	59.0%
2	Samsung Group	South Korea	49.29	150	76.7%
3	DowDuPont	United States	28.34	56	56.1%
4	Nippon Telegraph and Telephone Corporation	Japan	20.41	34	34.0%
5	Toyota Group	Japan	18.38	47	45.0%

*Date range for data is 2015–18



An adhesive, flexible wearable sensor measures heart and breathing rates.

topological materials, whose electronic structures impart unusual properties. Topological insulators, for instance, conduct electricity on their surface while blocking it from flowing through their bulk.

Google recently demonstrated Sycamore, a computer with 53 quantum bits, which performed a test calculation in just over three minutes that would take a supercomputer running at full capacity much longer (F. Arute *et al.* *Nature* **574**, 505–510; 2019). To really make a practical quantum computer, though, will take thousands or millions of quantum bits, and topological materials could create the technology to turn that into reality.

Additive manufacturing, also known as 3D printing, is another field that's taken hold in the past few years thanks to materials advances. When it was first developed, 3D printing was used mainly to make prototypes of various tools by having a laser transform powders into a solid, or extruding a thermal plastic through a nozzle. Researchers have since learnt how to work with a variety of materials, including metals and conductive plastics, to make parts for products. Boeing, for instance, is pursuing additive manufacturing as a more cost-effective way to make plane parts, and has invested in 3D printing companies. In 2018, the US aviation giant put money into Digital Alloys, a Massachusetts-based company with a technology that allows it to print different types of metals together to enhance their mechanical or thermal properties.

The multinational manufacturing conglomerate, 3M, which makes everything from transparent bandages to Post-It notes, has great interest in emerging materials innovations, says Greg Anderson, vice-president of its corporate research laboratory in Maplewood, Minnesota. But it usually asks whether an innovation can lead to a new platform, rather than a single product – more a 3D printing technology that can lead to a whole series of products, rather than the one widget made by 3D printing.

Despite the great potential for innovative products, there is significant risk that new materials discoveries will never amount to something marketable. Because of this, many companies like to support their development in small steps, by sponsoring academic research and funding university spinoffs, to see whether the technologies that emerge are worth more significant investment, Anderson explains. “We’ll do sponsored research to help us mitigate the risk,” he says. “We’ll spend a little bit of money in partnership with the university to advance the science and see how feasible that technology is.”

“In academia, we can dream something really big. What we do could be science fiction,” says Bao. “But some of those [ideas] could be turned into reality. That depends on the company and their creative thinking of how to use some of the crazy things that we invent.”

Neil Savage is a freelance science writer in Lowell, Massachusetts.

Wearing well

Stretch goals of a long-time partnership yield flexible electronics breakthroughs.

Stanford University's Zhenan Bao is taking wearable electronics to a radical new level inspired by skin — flexible, stretchable and unobtrusive. In August 2019, she and her team detailed a wearable sensor in the form of a sticker. It measures heart and breathing rates, and transmits this information via a flexible antenna and a radiofrequency identification tag to a receiver clipped to the wearer's clothes.

The device, described in *Nature Electronics* (S. Niu *et al.* *Nature Electron.* **2**, 361–368; 2019), is the latest milestone in Bao's long-running collaboration with the South Korean conglomerate, Samsung, and its in-house lab, the Samsung Advanced Institute of Technology in Suwon.

Bao's team had been working with Samsung on carbon nanotubes — cylindrical sheets of carbon atoms — for a decade when she pitched an idea for stretchable electronics in 2014. These were to be more than just rigid components with flexible joints, they were conceptualized as skin-like, intrinsically stretchable materials. This was a continuation of the previous work, as carbon nanotubes could be incorporated into such materials.

Samsung signed on. Its work with Bao and others has become a mainstay of the most productive international corporate-academic collaboration in the *Nature* Index. Bao describes the financial support as “substantial”. For decades Samsung has funded research at South Korean universities, and it acquired one of the nation's leading institutions outright, Sungkyunkwan University, which has campuses in Seoul and Suwon. Samsung sponsors research worldwide through its Global Research Outreach programme, open to all researchers.

In the early stages of the collaboration on stretchables, Samsung set performance targets for the electrical conductivity of the material that would be required to design viable products. Bao admits that she initially thought they were impossible, “but this kind of impossible target can make us more creative”, she says.

Mark Zastrow

People of substance

These early-career scientists are ushering in a world of new materials.

The five researchers profiled here are emerging as star competitors in a highly competitive field. Whether building better batteries or unlocking the brain's mysteries, they are deploying novel materials in novel ways to address pressing problems.

These three women and two men stood out from more than 10,000 researchers identified in the League of Scholars database as having five or more publications since 2015 in any of the following six materials-science-specific journals tracked by the Nature Index: *ACS Nano*, *Advanced Functional Materials*, *Advanced Materials*, *Nano Letters*, *Nature Materials* and *Nature Nanotechnology*.

Journals were selected for inclusion in the Nature Index based on perceptions of quality by independent panels of expert researchers. To be considered for selection, the materials scientists featured here had to be early in their careers, in this instance meaning it is less than five years since their first article citation. They have a strong track record of year-on-year publication and all have an outstandingly high annualized *h*-index, a metric which recognizes both the output and impact (using citations as a proxy for impact) across all journals while normalizing for differences in disciplines and career stages.

The citations data shown are drawn from the Dimensions for Digital Science database.

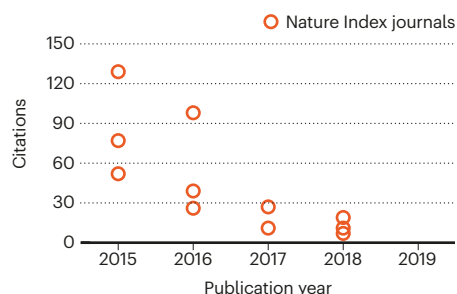
Lighting up the brain

Blanca del Rosal
Swinburne University of Technology

From a medical perspective, the greatest thing about synthetic nanomaterials, specifically nanoparticles, is their small size. Because they are on the same scale as human cell parts, they are useful as minute probes for spying on cellular machinery without causing too much disruption. Blanca del Rosal is developing a technology to do just that, using nanoparticle beams to unveil the inner workings of the brain. With her team at Swinburne University of Technology's Centre for Micro-Photonics in

Melbourne, Australia, she is figuring out how luminescent nanoparticles can track temperature changes during the treatment of brain tumours, which could lead to safer approaches.

Her research could also provide a minimally invasive way to observe how the brain behaves during strokes and comas. "The usual way to do it is by sticking a physical probe into the brain," says del Rosal.



In 2014, del Rosal tested the ability of luminescent nanoparticles to treat cancer in mice as a PhD student at the Autonomous University of Madrid in Spain. A nanoparticle solution injected into the tumours was heated with a laser to around 20% higher than normal body temperature. The heat generated by the irradiated nanoparticles reduced the size of the tumours, and the team used the fluorescence to monitor changes in temperature. In 20 days, the tumours in the mice were eradicated (E. Carrasco *et al. Adv. Funct. Mater.* **25**, 615–626; 2015). "It was the first time that anyone had used this contactless method for sensing temperature below the skin in animal models," says del Rosal.

Now a postdoctoral research fellow at Swinburne, del Rosal is investigating how to apply this new temperature-sensing approach to mapping brain activity in real time, with the hope of developing a non-invasive method of observing how the brain responds to conditions such as epileptic seizures, brain injuries and strokes.

She is also interested in how light-emitting nanoparticles can be used to assess the efficacy and safety of laser-based therapies used for restoring vision and treating



chronic pain. "There's still a lot we don't know about the brain, but light can give us more insight into how it works," says del Rosal.
Gemma Conroy

Stabilizing influence

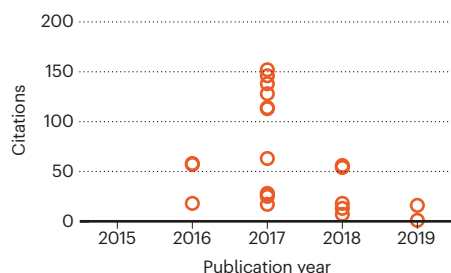
Emily Hitz
University of Maryland

An important aim for NASA is avoiding explosions during Moon missions. That's why Emily Hitz is investigating the lunar applications of solid state batteries as part of her three-year fellowship with the US space organization.



Blanca del Rosal is using nanoparticle beams to peer inside the brain.

Lithium-ion batteries, one of the most common types of rechargeable batteries on the market, use flammable liquid or gel electrolytes. Solid-state batteries, which use



solid electrodes and electrolytes made from ceramic, polymer or glass, offer far greater stability. “NASA is interested in having very safe batteries,” says Hitz.

Solid-state batteries are attracting interest as an emerging technology because of their safety, the large amount of power they can dispense per unit of battery size, and their long cycle life, meaning they can complete a high number of charge/discharge cycles at optimal capacity.

Based in the Department of Materials Science and Engineering at the University of Maryland (UMD), Hitz is interested in how new energy technologies can become

more accessible by incorporating cheap and easy-to-manufacture resources. “One day we might literally be using a block of wood to make a battery,” she says. “We use whatever we can to make it work. I like that flexibility.”

Having graduated from UMD with an electrical engineering degree in 2015, Hitz made the transition to materials science for her PhD. She wanted the opportunity to discover highly desirable material properties by having a stronger chemistry focus in her work. “I wanted to actually change the materials I was working with; to change their properties,” she says.

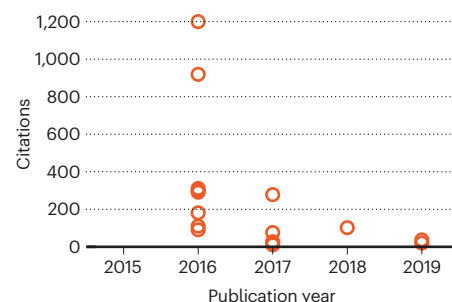
Hitz’s ultimate goal is to see new kinds of commercially viable batteries become part of a more diverse landscape of renewable energy technologies. “You can’t just sit in the lab and let the technology die there,” she says. **Bec Crew**

Solar cell optimizer

Juan-Pablo Correa-Baena

Georgia Institute of Technology

For materials scientist, Juan-Pablo Correa-Baena, the key to addressing the world’s growing energy demands is understanding solar cells at the nanoscale. At the Georgia Institute of Technology, he is uncovering the chemical reactions that underpin the performance of perovskite solar cells. They are cheap to manufacture and have shown remarkable increases in efficiency over the past decade, achieving a higher energy output than commercially available silicon solar cells.



Lead halide perovskites can convert sunlight into electrical energy, making them ideal materials for solar cells. Researchers have been experimenting with different ways to assemble perovskite solar cells to improve their stability and efficiency.

A major limitation of perovskite solar cells is that they don’t hold up well against moisture or extended periods of light or heat, so Correa-Baena is working on techniques to create a protective outer layer that doesn’t compromise the cell’s performance.

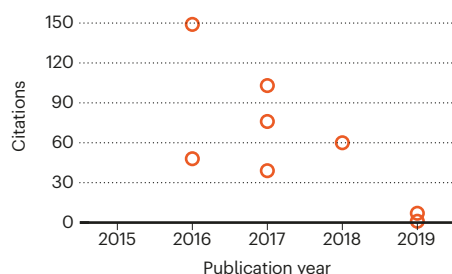
During his postdoc at the Massachusetts Institute of Technology (MIT) in 2018, Correa-Baena used the powerful light beams generated by synchrotron particle accelerators to probe the nanostructure of perovskite solar cells. He found that adding low concentrations of caesium and rubidium to perovskite solar cells can boost their performance. “It was one of the biggest discoveries I’ve had,” says Correa-Baena. “It opened up this whole new field of things that we could explore using synchrotrons.”

Improving access to power for rural communities in Correa-Baena’s home country, Colombia, is the driving force behind his research, he says. “There’s much more that we can do in research to try to push forward topics that are important to the world.” **Gemma Conroy**

Flexible thinking

Huang Yang
Shenzhen University

Huang Yang, an assistant professor of materials science at Shenzhen University in China, has emerged as a high-impact researcher through his work with supercapacitors, an alternative to batteries that could provide energy storage solutions for wearable and



bendable technologies. Supercapacitors are ideal in applications requiring many rapid charge and discharge cycles, such as electric cars, and could be the key to advancing foldable smartphones and television screens. In 2019, Motorola, Samsung and Huawei introduced bendable smartphones to the market, but although their screens can fold like a book, their lithium-ion batteries are not flexible. The current fix is to place either two electrodes of one battery, or two or more separate batteries, at either end of the screen, but this comes at a cost to the overall power of the device.

The ability to withstand bending, or ‘deformation’, is a major goal for energy storage technology. Huang is developing a flexible ‘shape memory’ supercapacitor that can be stretched and deformed. A heat trigger is then used to restore its shape and stability. The technology,



Huang Yang makes ‘shape memory’ supercapacitors, an alternative to batteries.

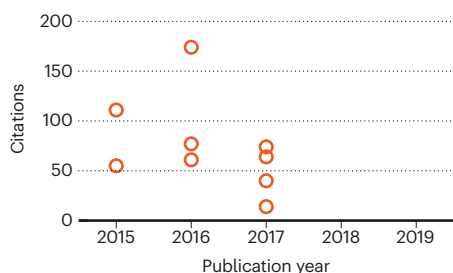
he says, would be ideal for powering wearable tech, as it can be sewn into fabric.

Huang and his colleagues take a multidisciplinary approach to developing supercapacitors. He says the unconventional way that they test the limits of the technology is revealing some promising new directions for flexible devices. “We’re combining organic and inorganic material, trying methods both at the nanoscale and macro-scale, and we’re attempting new and traditional materials,” he says. **Hepeng Jia**

Taking charge

Rachel Carter
US Naval Research Laboratory

An invention that won three researchers the 2019 Nobel Prize in Chemistry, lithium-ion batteries launched the portable electronics industry and are expected to play a crucial role in future renewable energy storage. Scientists are now working to solve problems relating to their safety, performance and the amount of waste they produce. In Australia, 2% of lithium-ion



batteries are recycled; in the United States and Europe, the rate is less than 5%.

Rachel Carter, a research engineer at the United States Naval Research Laboratory (NRL) in Washington, is investigating ways to expand the lifespan of lithium-ion batteries, while also pursuing alternatives, such as lithium sulfur batteries and sodium sulfur batteries.

The biggest challenge with rechargeable batteries, says Carter, is that the chemical reaction that generates power isn’t 100% reversible – meaning the battery loses capacity with every charge. “Inevitably, they will degrade,” says Carter.

“Our focus is on developing materials that produce high energy and high reversibility, but also to understand their behaviours from a materials science perspective, such as what happens when they’re heated or cooled.”

With her colleague, Corey Love, a materials research engineer at the NRL, Carter is finding ways to detect temperature variations, a common cause of dangerous leaks in lithium-ion batteries, without opening them.

“Using this technique to determine when thermal imbalances occur, we can detect unsafe cell conditions,” says Carter.

She is passionate about working with other young researchers on energy storage solutions. She is also leading initiatives within the NRL to foster professional development for female scientists. “I’m looking forward to seeing what happens with these battery chemistries, and collaborating with up-and-coming scientists is an exciting part of that,” she says.

Bec Crew

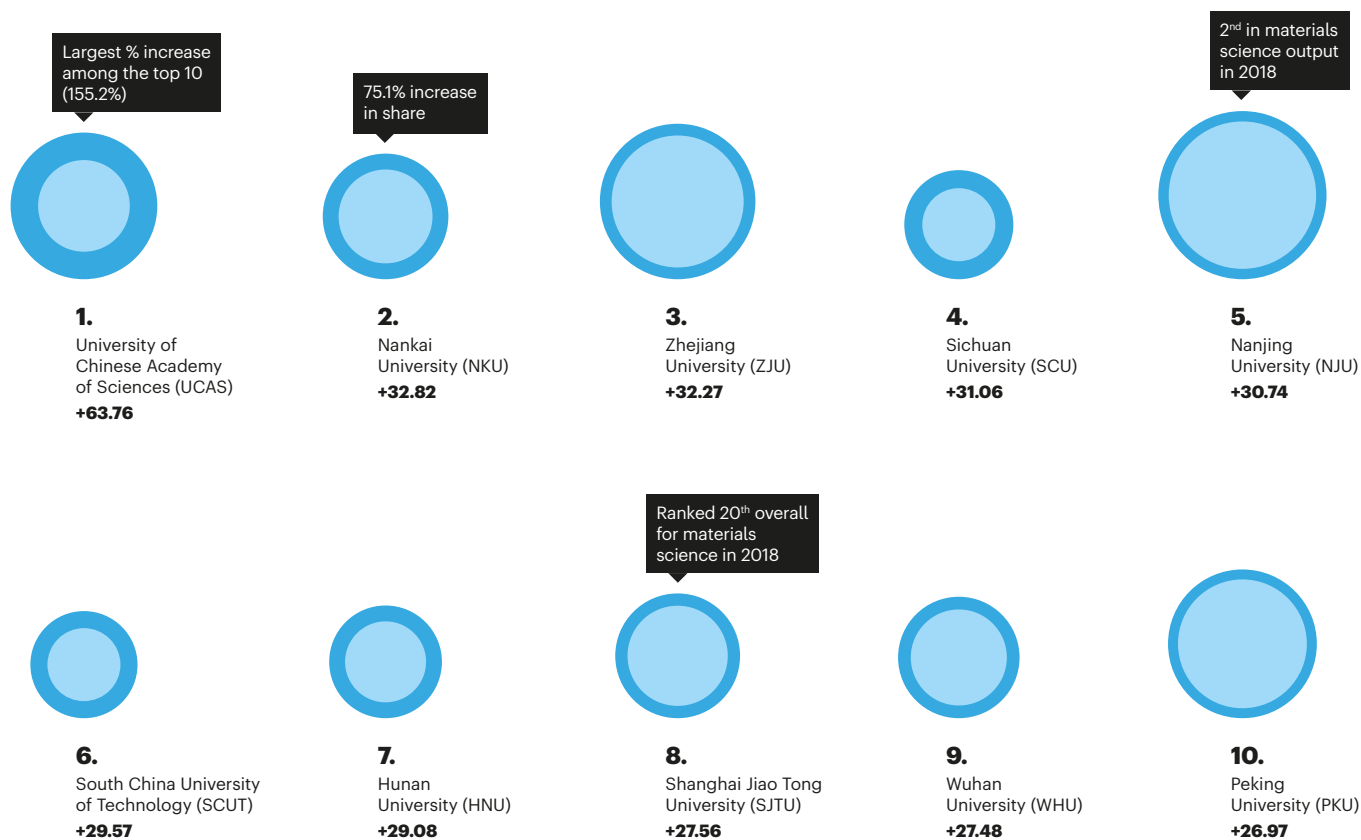
Scaling up

Chinese institutions monopolize the fastest-rising ranks for materials science output in the Nature Index.

Data analysis by Bo Wu. Infographic by Denis Mallet.

TOP 10 RISING INSTITUTIONS IN MATERIALS SCIENCE — GLOBAL

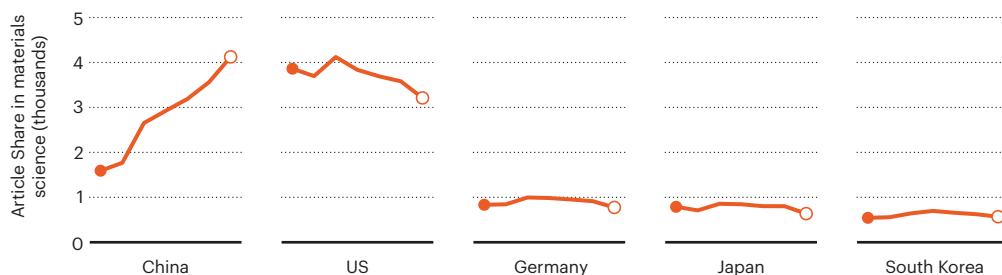
Institutions are ranked by the increase in their Nature Index Share in materials science between 2015 (inner circle) and 2018 (outer circle). The absolute change in their Share over the period is also shown.



TANGIBLE GAINS

In 2013, China's materials science Share in the Nature Index was less than half that of the US's, but China's Share surpassed the US's for the first time in 2018. Australia's Share grew by 42% in the six years to 2018, albeit from a low base. For 5 of the top 10 materials science countries (US, Germany, Japan, UK, and France) Share fell by between 0.9% (UK) and 19.8% (France).

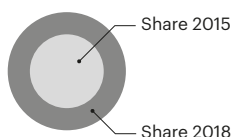
● 2012 ○ 2018



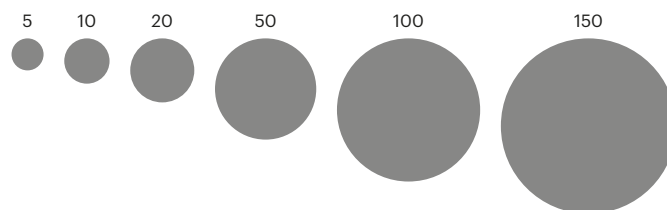
Legend

Institutions are ranked from left to right by the absolute increase in their Nature Index Share for materials science research in journals tracked by the index, 2015–2018. Circles are sized by an institution's Share in the index, which takes into account the proportion of their affiliated authors' contributions to articles published in the Nature Index journals. The absolute change in their Share over the period is also shown.

Growth in Share 2015–2018

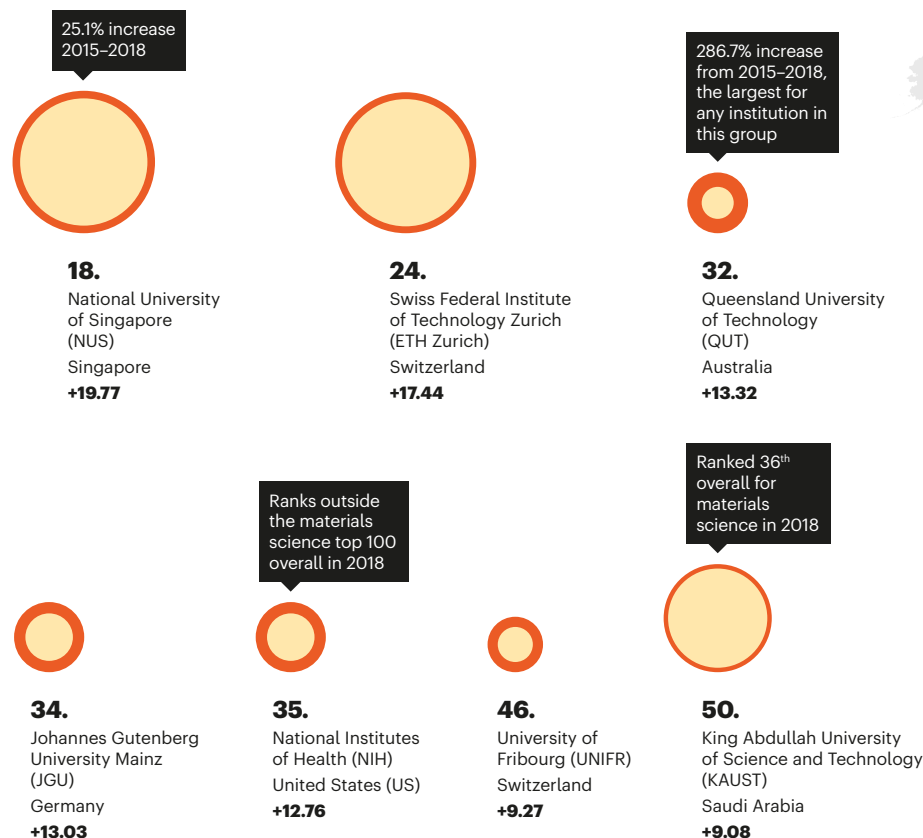


Institution's Share score by bubble size



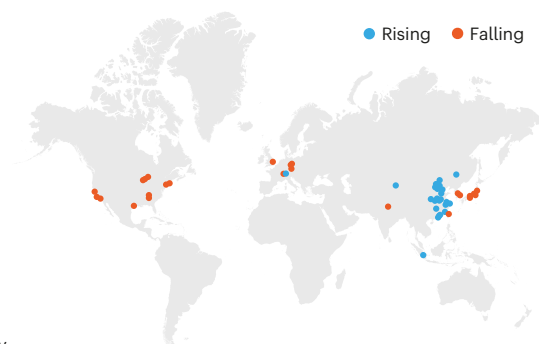
RISING INSTITUTIONS IN MATERIALS SCIENCE TOP 50 OUTSIDE CHINA

Just 7 institutions from outside China were ranked among the 50 institutions worldwide with the fastest-rising output in materials science 2015–2018.



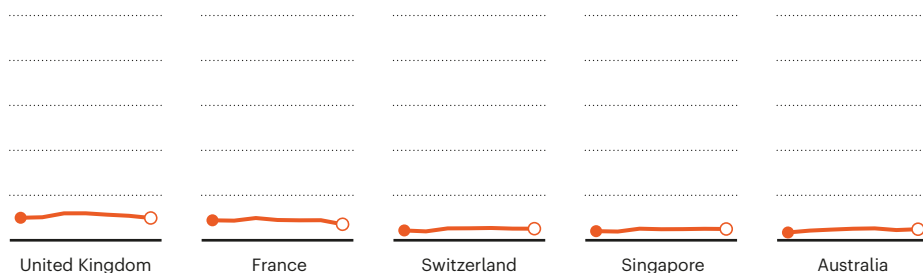
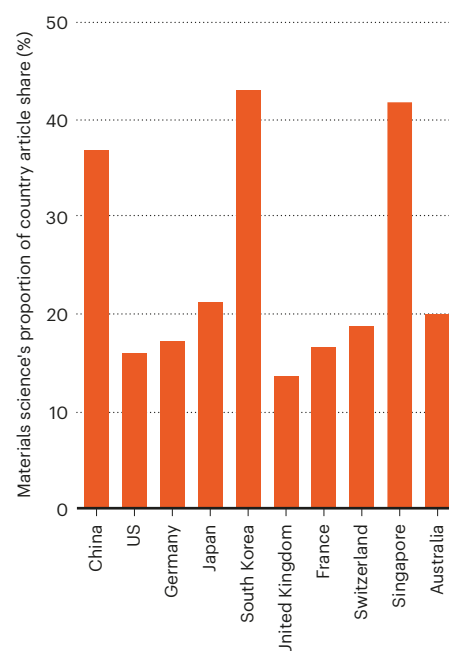
GLOBAL SHIFT

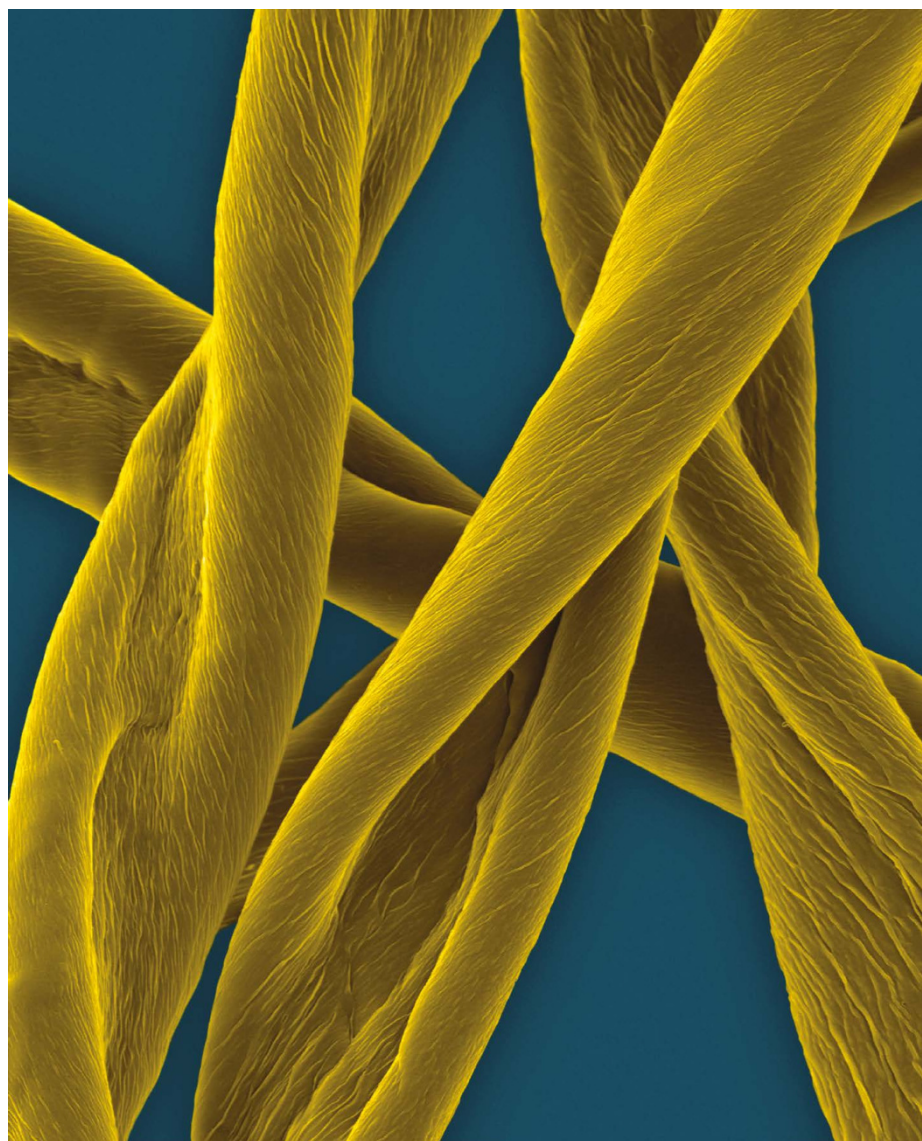
As institutions compete to author the finite number of articles which can appear in the 82 high-quality journals tracked by the Nature Index, the ascendancy of Chinese institutions translates to declines among some international competitors.



WEIGHT OF MATERIALS

Materials science Share as a proportion of total Share in the Nature Index is shown for the top ten countries, ranked left to right by their 2018 materials science Share.





Cotton fibres can become conduits for ions, atoms or molecules, and generate electricity.

Energy harvesters pick up power

New technologies are channelling incidental energy into practical uses. **By Mark Zastrow**

Wood seems an unlikely material for the forefront of an energy revolution. But materials scientist Tian Li believes it could unlock electricity from a source of energy that is both sustainable and ubiquitous: our own body heat.

The key lies in the remarkable fibrous structure of cellulose, the main component of wood, paper and cotton. These nanoscale fibres are “a beautiful structure” and a marvel of natural molecular engineering, says Li, a postdoctoral researcher at the University of Maryland. When treated with a chemical bath,

the fibres can become conduits for ions, atoms or molecules with a net electrical charge. By exploiting how ions move in the presence of heat, they can generate electricity, even from small temperature differences, recovering energy that would otherwise be lost to the atmosphere.

This is just one way that researchers and companies are trying to scavenge energy from our ambient environment in its many forms, including radio waves, vibrations and sound. The global ‘energy harvesting’ industry is worth around US\$500 million, and growing by roughly 10% each year, according to industry reports. Its development is seen as crucial to the expansion of the Internet of Things (IoT), a vision in which our homes, offices and infrastructure are made ‘smart’ with vast numbers of networked sensors. Proponents also say it can contribute to a greener world. “We need as many technologies as possible to generate sustainable energy,” says Li.

Wooden heat

Li and her colleagues at the University of Maryland developed a process for soaking slabs of common American basswood in a chemical bath of sodium hydroxide (T. Li *et al. Nature Mater.* **18**, 608–613; 2019). This extracts the lignin, a natural polymer intertwined around the cellulose fibres, which gives wood its brownish hue. The treatment also breaks the hydrogen bonds along the remaining cellulose nanofibres and forms a crystalline structure with highly aligned molecular strands, along which sodium and hydroxide ions can freely travel.

Li fabricated similar materials when she studied semiconductors in graduate school, but this process is much simpler. “I recall having to go to the fabrication lab for a week to make sure all the channels were aligned and perfect,” she recalls. “Then, when I did my postdoc, I was looking at this material [wood] and was amazed. It had perfectly aligned channels.”

When placed against a heat source, the nanostructure of the cellulose allows the sodium ions to drift away from the source more easily than the hydrogen ions, creating an electric potential that can later be tapped to release electricity. This treated cellulose can harvest heat energy from sources just 5 °C warmer than its surroundings.

Li envisions shirts made of similarly treated cotton, which is nearly pure cellulose, that could harvest body heat, storing it in a battery to recharge phones, or charging them wirelessly in pockets. “Previously, when people talked about wearable flexible devices, they put this device on top of a substrate,” she says. But in this case, a separate surface is not needed. “Your cotton t-shirt can be a device itself.”

DENNIS KUNKEL MICROSCOPY/SCIENCE PHOTO LIBRARY

Wood could also potentially tap into other industrial sources. It is estimated that the global amount of low-grade wasted heat from industrial processes and other sources is enough to power more than 6,000 typical one-gigawatt nuclear reactors.

Sensing an opportunity

Over the past decade, the potential of energy harvesting has been alternately hyped and challenged. Some applications have been stymied by physics: absorbing a sufficient amount of energy requires devices that are too large to fit in small sensors that can use cheaper batteries instead.

“The key is coupling the energy harvester to a low-power processor or radio,” says Mike Demler, a semiconductor technology analyst with the Linley Group, a Californian consulting firm.

Many people view the sensors in IoT devices as a key application for energy harvesting. “Sensors usually consume very little power,” says Kim Dong In, an electrical engineer at Sungkyunkwan University in Suwon, South Korea. The problem with IoT sensors, he says, is the vast number of them that need to be deployed, and the challenge of recharging them or replacing their batteries. Energy harvesting is a natural fit for powering sensors that are not easily accessible – behind the walls of a smart home, for instance, or embedded in the structure of a bridge to monitor its integrity. Kim sees these as perfect environments for devices that can recharge themselves by harvesting energy from radio waves. “We cannot see them, but a lot of energy is available from surrounding electromagnetic signals such as TV, wifi and cellular phone towers,” he says.

In such rechargeable devices, an antenna picks up radio waves, which induce an alternating current. This is passed along to a diode, which converts the signal into a direct current that can charge a capacitor. The challenge is that the effective charging range is only a few metres and diodes are not very efficient. For example, the signal strength needed to generate useful power far exceeds the signal strength a wifi router needs to maintain a strong connection.

Kim proposes a solution: using a separate base station designed purely to transmit power, not information. In a smart home, this could be incorporated into a robotic vacuum cleaner that roams the house, recharging sensors as it goes. For a bridge, a drone equipped with a power transmitter could fly down its length to top up embedded sensors. This, Kim says, makes radio-frequency harvesting one of the more flexible energy harvesting techniques.

The concept could be taken even further by



An illuminated installation that takes its energy from the footsteps of people passing by.

the use of wallpaper with carefully designed metasurfaces – surfaces with nanoscale structures made of individual units, akin to pixels, that can manipulate the wavefronts of light. By applying a small voltage to each unit, researchers can create an artificial lens, focusing and steering beams of radio waves to their targets.

Some energy harvesting techniques have more proven track records. Wearable devices

“We need as many technologies as possible to generate sustainable energy.”

that harvest energy have been around for centuries: automatic watches, which wind themselves using an oscillating weight as the user moves them, date back to the 1770s. The first electronic device to harvest its own energy in this way was also a watch, patented in 1988 by Seiko. Solar cells have powered watches and calculators for decades, and are frequently used in IoT sensors, says Demler.

Another relatively well-established energy harvesting technique is based on piezoelectrics, materials that generate electric charge when squeezed or pressed. In 1998, MIT

researchers were some of the first to experiment with embedding them in shoes – a technique that garnered the interest of the US military. By 2008, Japan was experimenting with embedding piezoelectrics in the floors of subway stations to power turnstiles.

Work continues on a variety of experimental techniques. One promising class of devices works by generating triboelectricity, also known as static electricity. These small devices, called triboelectric nanogenerators, use friction between their moving parts, such as two sheets of flexible polymers or a silicone rubber ball that bounces around a hollow plastic sphere. The first such devices were developed by a team at the Georgia Institute of Technology in Atlanta.

Ultimately, successful implementation of energy harvesting comes down to tailoring a technique for a specific use case, says Demler. Further advances in computer chips that can use lower amounts of power will help, too. As for Li’s nano-engineered wood, she says she’s yet to start work on commercializing it, but plans to once she lands a faculty position. “It’s such an open field,” she says. “And I feel this is truly a field worth exploring.”

Mark Zastrow is a science writer in Seoul, South Korea.